

Reflections on scientific fraud

An exhaustive investigation has revealed significant contamination of the physics literature by a researcher. Such incidents are difficult to prevent, but all involved can try harder.

Long gone are the days when scientific frauds could be dismissed as the work of the mad rather than the bad. The unhappily extensive record of misconduct suggests that many fraudsters believe their faked results, so attempts at replication by others represent no perceived threat. Just what was going through the mind of the condensed-matter physicist Jan Hendrik Schön has not been revealed by the inquiry into his work, which last week reported 16 incidences of fraud (see page 419). Was he trapped in a spiral of expectations that could only be fulfilled by deception?

Maybe time will tell. In the meantime, others need to look to their laurels. Co-authors have been exonerated of fraud, but those who had most experience of the materials being explored will be searching their consciences over what they might have been able to prevent had more self-critical attitudes been adopted. Editors, referees and co-authors: be cautious if one co-author has been single-handedly and prolifically ground-breaking without replication by their colleagues or others.

The report reaches no conclusions about the role of the journals, including *Nature*, *Science* and *Applied Physics Letters*. *Nature's* editors have looked at their files, including the timing and the state of the concurrent literature, and the scatter of Schön's research across different types of material. *Nature's* peer-review processes have picked up fraud on occasion, and if a referee of Schön's *Nature* submissions had looked at papers on different materials, he or she might have spotted the duplications in data that, in the end, were the smoking gun. But that would have been by luck, rather than by reasonable expectation.

In some media reports, journalists and a few scientists who are unconnected with the Schön investigations have taken the opportunity to make potentially damaging assertions about journals, including *Nature*: that in order to compete or to publish exciting results, journals will cut corners in peer review, overrule hostile reviewers or select sympathetic ones.

We at *Nature* unequivocally reject such charges. The publication history and files of these particular papers and the editorial policies

and interests of *Nature* are completely at odds with these assertions. *Nature* has nothing to gain by the pursuit of glamour at the expense of scientific quality, considering, not least, the criticisms, corrections and retractions we would then habitually be forced to publish. There is more than enough rock-solid and splendid science to publish. Furthermore, it is a strict policy of *Nature* that our Letters and Articles are selected for their outstanding scientific impact, sometimes also taking into account relevance to public policy issues, but never simply because the results will make headlines.

Reviewers are not selected so as to ensure a positive outcome for exciting work. Preference is given to experts with a proven track record in providing thorough and critical advice. There is always the possibility that exciting papers from groups with an excellent track record will be viewed less critically than those from an unknown quantity, as referees and editors might take more on trust, and it is here that journal reviewers and editors may find the investigating panel's report to be salutary (see www.lucent.com/news_events/researchreview.html).

Nature stands by its insistence that all co-authors should have seen and agreed to a submitted paper. We have indicated on the online versions of the five *Nature* papers that they are contaminated — see also page 425. We have invited all co-authors whom we could contact to send in retractions. As we have done in the past, we are prepared to publish retractions by a majority of authors in the face of non-cooperation by one co-author, making any author's abstention clear to readers. As *Nature* goes to press, the signs are that Schön's co-authors will agree to such retractions.

The Schön scandal is generating widespread attention and critical introspection. Such misconduct has occurred before, mostly in the life sciences, and the scientific community has improved its procedures for investigating misconduct when it arises. Moreover, in many countries, government agencies have introduced principles of laboratory management in order to minimize the potential for fraud. A key question now is whether such guidelines are being implemented. ■

Time to eradicate malaria

The fight against the disease needs not only money but also coherence.

Malaria kills over one million people every year, mainly in Africa south of the Sahara. Another 300 million to 500 million have the disease. But despite the dedication of many fighting this formidable enemy, progress is chronically slow. Politicians and research administrators are prone to throwing in the towel. Basic researchers often care more about their next grant proposal or publication than going that extra mile to tailor their work better to meet the needs of researchers in poor countries for applicable tools. And bureaucrats in the international health system (see page 422) sometimes pay more attention to, well, their bureaucracy.

The time is right for a sustained effort to eradicate malaria. This week sees the publication of the genomes of *Plasmodium falciparum*, the deadliest malaria parasite, and its proteome (see pages 498–542), and of *Anopheles gambiae* (*Science* **298**, 129–149; 2002), the mosquito

vector. These, combined with the human genome, mean that the scientific infrastructure for a complete understanding of the biology is now in place, as well as for generating drug and vaccine targets in a faster and more rational manner (see also pages 426 and 429).

Resources for malaria research and control are still scandalously scarce, but are getting better for control, thanks to the Global Fund to Fight AIDS, Tuberculosis and Malaria, whose US\$100-million input this year will double the overall funding for control. To capitalize on the science, a comparable Global Health Research Fund is also required. Most importantly, there is a need for everyone involved to coordinate their efforts in basic genomics, drug and vaccine discovery, and implementation in the field. There is a rallying call for the international health and research community that may sound trite but is both apt and timely: "Together, we can and must stop malaria." ■





Double trouble
Is there more than
one Saddam Hussein
out there?
p421



Poor health
Global anti-malaria
programme needs
reorganization
p422



People power
Residents resist plan
for Rocky Mountain
bioweapons lab
p423



Last in line
Bactrian camel on
UN's most
endangered list
p425

Misconduct finding at Bell Labs shakes physics community

Geoff Brumfiel, Washington

Physicists are coming to terms this week with one of the most audacious scientific frauds ever uncovered. The research in question was carried out at one of the world's best-known industrial laboratories and published in top journals — including this one (see editorial statement on page 425).

An independent committee charged with reviewing work done by Jan Hendrik Schön at Bell Laboratories in Murray Hill, New

Jersey, released its findings on 25 September. It found a “preponderance of evidence” that he falsified or fabricated data in 16 of the 24 alleged cases of misconduct that it looked at, involving 25 published research papers.

“Hendrik Schön showed reckless disregard for the sanctity of data in the value system of science,” said the report of the committee, which was chaired by Malcolm Beasley, an electrical engineer at Stanford University in California. Bell Labs — part of telecommunications corporation Lucent — fired Schön the evening before the findings were released. But the affair is far from over, with physicists asking not only what could have been done to prevent the fraud, but also what will become of more than 100 papers that Schön authored since arriving at Bell from Germany in 1998.

In a statement accompanying the report, Schön admitted to making “mistakes”, but said that he “never wanted to mislead anybody or to misuse anybody’s trust”. He added: “I truly believe that the reported scientific effects are real, exciting, and worth working for.” He was unavailable for further comment.

Schön’s work to create transistors out of



Bell Labs: reinforcing in-house peer review in the wake of scientific misconduct verdict.

single molecules and induce superconductivity in carbon ‘buckyballs’ won him acclaim, and his findings adorned the covers of *Nature* and *Science*. But researchers struggled to replicate his results, and as Schön’s publication list grew, so did the community’s scrutiny. In May, a group of researchers informed Bell Labs that they had discovered a series of three graphs that



Jan Hendrik Schön: fired from Bell Labs after an inquiry found falsification of data.

According to the Beasley report, Schön used many different techniques to fabricate and falsify data for his papers. These ranged from the use of mathematical formulae instead of data to plot points on a graph, to reusing data from one experiment in plots associated with another one. In addition, says the report, he often changed his account of how data were collected, when questioned by investigators.

One experiment used large electric fields to induce superconductivity in carbon buckyballs (J. H. Schön, Ch. Kloc & B. Batlogg *Nature* 408, 549; 2000). The report catalogues how Schön used a mathematical function to create a graph of the buckyball’s electrical resistance as a function of temperature.

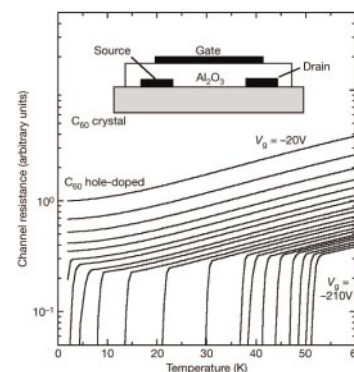
“Figure 32 [Fig. 1 of the paper (right)] shows an amazing systematic variation of

the resistance versus temperature,” the report says. “Even the deviations are reproduced with high precision between the various curves... It is clear that these are not real data: they have been generated using a mathematical function.”

Initially, Schön described how the apparent smoothness of the curve could result from smoothing of real, noisy data points. But, when faced with the reality that such smoothing could not explain the extraordinary fit, he acknowledged that the data were in fact spliced-in analytic functions. “I thought that a smoother curve would look much better,” he told the committee.

After admitting this falsification, Schön supplied a plot showing more realistic curves, but not the original data. Schön told the panel that he had deleted all his primary

Keeping up appearances



data because his computer at Bell Labs was short of memory.

♦ http://www.lucint.com/news_events/researchreview.html

G.B.

appeared identical in different publications, down to what should be random noise (see *Nature* **417**, 367; 2002). This revelation, together with additional suspect graphs, led Bell to convene Beasley's panel.

"The report shows that scientific misconduct occurred," says Beasley. "But there are other questions to ask." The panel also took a close look at Schön's most frequent

co-authors: Christian Kloc, Zhenan Bao and his supervisor Bertram Batlogg. "We found it very difficult to deal with the issue of the responsibility of co-authors," says Beasley. The committee cleared Kloc and Bao of any wrongdoing, but it raised questions about Batlogg, who left Bell in autumn 2000 for the Swiss Federal Institute of Technology in Zurich, but whose name continued to appear on Schön's papers throughout 2001.

Batlogg: regrets in retrospect.

The report asks if Batlogg should have "insisted on an *exceptional* degree of validation of the data in anticipation of the scrutiny that a senior scientist knows such *extraordinary* results would surely receive" or even "crossed the line of trust and questioned the integrity of the data". In the end, it says: "The Committee does not consider itself qualified to make a specific judgement in this case."

Some researchers close to the case claim that some responsibility must rest with the co-authors, particularly Batlogg. "You can't just take the kudos of authorship and then say, 'It's not my fault,'" says Charles Lieber, a chemist at Harvard University in Cambridge, Massachusetts, who was one of the first to note problems with Schön's data. "I think the buck stopped at Batlogg," asserts Arthur Ramirez of Los Alamos National Laboratory in New Mexico, who has devoted his own research to replicating one of Schön's results.

"As co-author I acknowledge a responsibility to ensure the validity of data in publications," Batlogg said in an e-mail. "I have learned, with the deepest of regrets, that the verification measures I have followed in this extraordinary case were not adequate to prevent or uncover scientific misconduct. I have placed, in retrospect, too much trust in my highly talented collaborator."

Beasley: headed investigation.

Bao and Kloc prepared materials for Schön's experiments and are not regarded as sharing responsibility for his findings.

Researchers are now asking if Bell Labs management should have had a tighter grip on what its staff were doing. "This has to be a management failure on some level," says Nobel laureate Philip Anderson of Princeton University in New Jersey, a former Bell researcher. Cherry Murray, vice-president of physical sciences at Bell, says that the laboratory responded promptly to two complaints it received about Schön's work, in October 2001 and February 2002. "In both cases Schön was very apologetic and the explanations did seem to reflect honest mistakes," Murray says, "but with 20:20 hindsight, these cases should have alerted us to fraud." She adds that steps have been taken to reinforce peer review inside the lab — including the creation of an internal server where all papers must be posted before submission for external publication.

Others ask why editors at journals, such as *Nature* and *Science*, weren't more suspicious of Schön's astonishingly high publication rate. Some critics charge that journals ignored negative feedback from the community because they wanted to publish high-

profile work. "*Nature's* editorial and refereeing policy seems to be influenced by the newsworthiness of the work, not necessarily its quality," says Anderson. "And *Science* seems to be caught up in a similar syndrome."

But Karl Ziemelis, physical sciences editor at *Nature*, says that there was no evidence of wrongdoing when the papers were accepted for publication. "On several papers, referees had questions over the interpretation of the data," Ziemelis says. "But they were nearly unanimous on the importance of the findings" (see Opinion, page 417).

"You can't design a peer-review system that's an absolute guarantee against clever fraud, but obviously when something like this happens, you're going to look at your practices," adds Donald Kennedy, *Science's* editor-in-chief.

Journal publishers now have to decide what to do about the status not just of the 25 papers reviewed by the Beasley panel, but also of the rest of Schön's published work. Kennedy and Ziemelis both say that their journals will request retractions of the papers named as fraudulent by the committee. *Science* had not decided what to do about the other work, but the co-authors on all Schön's *Nature* papers will be asked if they want to retract.

Rising star crashes back to Earth

Alison Abbott, Munich

Jan Hendrik Schön (right), who turned 32 in August, was truly the Wunderkind of German physics.

In a society where young scientists often struggle to establish themselves, Schön's extraordinarily successful track record gave him almost legendary status. And until doubts arose about the integrity of his work earlier this year, he was poised to return to his native land in glory — as the youngest ever co-director of one of the prestigious Max Planck Institutes.

Schön's meteoric career began at the University of Constance, where he studied physics as an undergraduate and then completed a PhD in 1997. In Constance he is remembered not just as a brilliant student, but as a self-effacing young man, well liked by colleagues, whose research was seen as diligent and solid.

His PhD topic was the development of semiconductor systems for use as solar cells. He tried to develop a junction between two copper selenide semiconducting materials that would convert light into electric current. He failed to build the junction — other groups later also failed to get the same system to work — but his well argued thesis won him a strong recommendation for a postdoctoral position at Bell Laboratories, supported by an 18-month grant from



Germany's main grant agency, the DFG.

Schön impressed colleagues at Bell Labs, and was offered a permanent position. For a few months at the beginning of 2001 visa problems sent him back to Constance, where he continued working on his Bell Lab projects. It was a time of breathtaking productivity: Schön wrote a paper every

Ramirez says that he now questions everything Schön has published. Colleagues have passed him eight more suspicious figures, including one that appeared in a *Nature* article not previously implicated (J. H. Schön, Ch. Kloc and B. Batlogg *Nature* **406**, 702; 2000). Lieber says that “any paper Schön was first author on since arriving at Bell Labs should be withdrawn”.

Along with many other physicists, Myriam Sarachik of the City University of New York, president-elect of the American Physical Society (APS), believes that science is self-correcting, and that Schön's fraudulent work will eventually be either ignored or discarded. “The scientific method does ultimately weed out this sort of thing,” she says.

Nevertheless, Sarachik says, the APS will review its misconduct policy in light of the Schön case and another high-profile fraud case involving the claimed discovery of element 118 (see *Nature* **418**, 261; 2002). The society may also develop programmes to teach graduate students good laboratory practice, and designate an ombudsman to receive complaints of misconduct. “These latest episodes have made us realize that we'd better look at how we conduct our business,” Sarachik says. ■

eight days on average during 2001, according to the Beasley report.

Schön's overnight success prompted the Max Planck Society to create a co-directorship especially for him at the Max Planck Institute for Solid State Research in Stuttgart. Eighteen months ago it initiated the lengthy *Berufung* process, a deliberation over a candidate's suitability that normally precedes a formal job offer.

“The procedure was, as usual, very careful and involved a lot of evaluators, none of whom raised objections,” says the institute's managing director, Martin Jansen. “During interviews, Schön came over as a very kind and impressive personality who presented his work very convincingly.”

Schön had signed no agreement with the Max Planck Society, but negotiations over his employment package had begun when suspicions about his work surfaced publicly. Negotiations were then put on ice, and Jansen has now recommended their cancellation.

Meanwhile, the DFG is studying the report's findings to see if the work that it supported involved misconduct. If so, says a spokeswoman, the DFG will immediately embark on its own inquiry. Sanctions for misuse of DFG funds range from a public warning to a demand for the return of grant funding. ■

Saddam gives viewers double vision

Quirin Schiermeier, Munich

Rumours that the Iraqi dictator Saddam Hussein uses doubles have been endorsed by a German face-recognition expert, who says he has evidence that television pictures of Saddam are really of his doppelgängers.

Commissioned by the German public TV network ZDF, Dieter Buhmann, a forensic scientist at Saarland University in Homburg, analysed 450 photographs of Saddam from the ZDF's archives and found significant deviations. “There is clear evidence that Saddam has used doubles in all but one appearance in public since 1998,” he says.

Buhmann used an unpublished computer method that he first presented at the 2000



meeting of the International Association for Craniofacial Identification in Washington.

Chris Solomon, a biometricist at the University of Kent, says the results are plausible, but should be interpreted with caution. “You can never be totally sure whether or not two people are identical from the basis of photographs,” he says. ■

Former ImClone chief faces allegations of past misdeeds

Jonathan Knight, San Francisco

A biotechnology entrepreneur who is facing high-profile charges of insider trading on the US stock market was repeatedly fired for suspected misconduct during his research career, according to allegations published by *The Wall Street Journal*.

A front-page story in the US newspaper on 27 September claims that Samuel Waksal, the former chief of biotech firm ImClone Systems in New York, was dismissed over questions of misconduct on four occasions in the 1970s and 1980s. Each time he quickly landed another research job at a prestigious institution, the story says. Scott Tagliarino, a spokesman for Waksal at Rubenstein Associates in New York, says that Waksal had no comment to make on the allegations.

But if true, the story highlights the difficulties faced by US institutions in trying to police scientific misconduct. University lawyers often prohibit faculty members from telling a suspect's prospective new employer that a misconduct investigation is under way, experts in scientific conduct say. And even if a scientist is found guilty, the university may keep quiet to avoid being sued for libel, says David Goodstein, a physicist and vice-provost of the California Institute of Technology who developed its policy on scientific misconduct.

Waksal is currently under investigation for allegedly informing friends, including homemaker celebrity Martha Stewart, of the impending rejection by federal regulators of ImClone's star cancer drug Erbitux, allowing them to cash out their stock.

According to the *Journal* article, he was ejected from a lab at Stanford University in 1974 on suspicion of lying about obtaining certain reagents. His employer at the time,

biologist Leonard Herzenberg, confirmed this account in an interview on 30 September, adding that he told Waksal's next employer of his concerns by telephone with Waksal on the line — but that Waksal denied having lied and the warning went unheeded. “He was a real charmer,” Herzenberger says. “He charmed everybody; people believed him.”

But the problems continued, the *Journal* alleges. Waksal lost subsequent positions amid concerns about misconduct, it says, at the National Cancer Institute in 1977, at Tufts University School of Medicine in Boston in 1982, and at the Mount Sinai School of Medicine in New York in 1985 — the year he founded ImClone. Henry Wortis, who collaborated with Waksal at Tufts, told *Nature* that he heard about the allegations of misconduct from several colleagues during the 1980s, after Waksal failed to deliver materials promised in a collaboration. ■



Samuel Waksal: questions raised over his career.

ZDF

D. COOK/AP

Weak leadership threatens anti-malaria drive

Declan Butler, Paris

A major global initiative that aims to halve the world's malaria burden by 2010 will fail unless it is speedily revamped, according to the scheme's first major external evaluation.

A panel chaired by Richard Feachem, director of the Institute for Global Health in California, says that the project, Roll Back Malaria (RBM), needs stronger leadership, a management overhaul and tight focus on a small number of countries where its ideas stand a chance of being put into practice.

The RBM initiative was launched in 1998 by the World Health Organization (WHO), as a partnership with the World Bank and the United Nations, and has since been joined by more than 90 other partners. It aims to direct a comprehensive global effort against the disease, which affects up to 500 million people and kills a million of them each year.

The evaluation was made by seven experts in research, health and economics. It found RBM successful in advocacy and in building a consensus about priorities: after languishing for decades, malaria research and control is high on political and scientific agendas, it says. Funding has almost doubled since 1998 to around \$200 million in 2002, \$35 million of which is channelled directly through RBM (see News Feature, page 426).

At a summit in April 2000, 44 of Africa's 50 malaria-affected countries pledged to



Sick of waiting: slow progress against malaria is failing sufferers such as these patients in Kenya.

support RBM's main goals — particularly that, by 2005, 60% of sufferers get immediate access to treatment. But progress has been sporadic. Only a handful of countries have increased funding and staff to anywhere near the required levels. RBM's own projects, the evaluation says, have too often been implemented in isolation from the broader health policies of recipient nations.

The report adds that RBM and its partners have given inadequate, even conflicting, technical advice to governments. It recommends the creation of technical support

networks to collate the best advice, and of an independent governance board to which the RBM secretariat should be accountable. It calls on RBM to reorganize its efforts, focusing on 8–12 countries that have the political commitment to enable rapid progress.

David Heymann, executive director of WHO's communicable diseases division, says the report agrees with an internal review done last year. "The external evaluation has been very useful," he says. "By the end of this year we expect to have strengthened the partnership based on its recommendations." ■

Criminal courts 'should take genetics into account'

David Adam, London

Stephen Mobley's defence at his trial for killing a pizza-store manager in Jackson, Georgia, was unusual. His lawyers had pointed to a family history of violence, and argued for leniency on the basis that genetic factors contributed to his crime.

Mobley's defence was rejected, and he is now on death row. But in a controversial study published this week — "Genetics and Human Behaviour: The Ethical Context" — a leading bioethics group says it may only be a matter of time before courts use genetic evidence to help determine punishments.

The London-based Nuffield Council on Bioethics says that genetic factors should sometimes be cited in mitigation for a crime, just as social factors, such as being a victim of child abuse, already are.

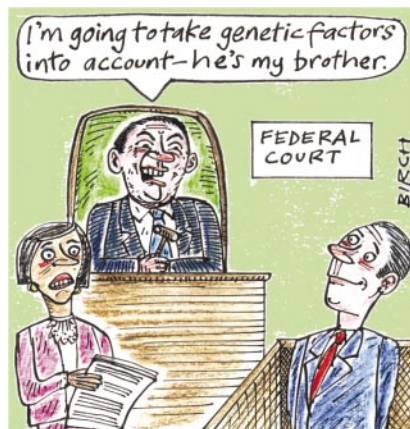
"If it is legitimate for a judge to take upbringing into account then it's legitimate to take genetics into account, providing the data are good enough," says Martin Bobrow, who is head of the department of medical genetics at the University of Cambridge, and deputy chairman of the Nuffield Council.

Some researchers do claim links between a person's genes and antisocial behaviour. A certain form of a gene that breaks down neurotransmitter chemicals has been found to make men more likely to be violent, if they were maltreated as children (*Science* 297, 851–854; 2002). Before genetics can be used in courts, however, Bobrow says that such results would have to be successfully

replicated and the magnitude of the effect quantified. For some associations this could be done "within a decade", claims Andrew Wilkie, a medical-genetics researcher at the University of Oxford and a member of the working party that prepared the Nuffield Council report. He adds, however, that very few people would be affected.

But not everyone agrees that genetic influences should be used this way. Sceptics include Craig Venter, the former boss of Maryland-based Celera Genomics. Using genetics as mitigation in courts would be a "dangerous leap", Venter said last week. "I don't see anything in the genetic code that forgives criminal activity," he says.

Some also fear that using genetics in sentencing could lead to the use of gene therapy to try to 'cure' criminals, and the genetic screening of populations for likely offenders. "This poses some fundamental questions for the legal system," says Margaret Somerville, director of the Centre for Medicine, Ethics and Law at McGill University in Montreal, Canada, "because the technology will surely improve." ■



Residents force review of biodefence lab

Rex Dalton, San Diego

Officials charged with distributing billions of dollars for US biodefence research are keeping a nervous eye on a small town in Montana, where residents have delayed a project to build a new pathogen-research laboratory.

A summer of discontent in Hamilton led the town's Rocky Mountain Laboratories to order a lengthy environmental assessment of its plan to build a Biosafety Level Four (BSL-4) complex. BSL-4 labs are licensed to handle very contagious agents, such as ebola and smallpox. Environmental consultants were recruited last month to do the study, and the process will add 12 months to the four-year construction schedule.

The decision to build a BSL-4 lab at the Rocky Mountain centre was one of the first made by the National Institute of Allergy and Infectious Diseases (NIAID) after it was provisionally awarded \$4 billion for biodefence research in February. Of this, \$430 million was for building secure facilities. The specific pathogens to be studied have not been confirmed, but the Hamilton lab is considered a suitable venue as it already has BSL-3 facilities for studying diseases transmitted by animals. Last year's anthrax attacks focused attention on the scarcity of US facilities for biodefence research. The only large BSL-4 labs, for example, were at the Centers for Disease Control and Prevention in Atlanta, Georgia, and at a US Army base in Fort Detrick, Maryland.

Problems with consulting the 4,000 residents of Hamilton were evident early on. Lab administrators were criticized for holding the first public meeting on the evening of St Valentine's Day. At a second meeting in July, residents were told that they had just two weeks to submit comments for consideration in an environmental assessment of the lab.

"That was ludicrous," says Mary Wulff, a former policewoman who helped to found a coalition to oppose the lab's development. "Things started to snowball after that. Residents are walking around the town saying 'why don't they just put in the middle of nowhere?'" The coalition complained repeatedly to the lab, citing concerns such as the inability of the region's volunteer fire and emergency teams to deal with injuries resulting from an escape of pathogens.

Faced with increasing anger from the local community, lab officials were forced to agree to a more detailed study of the environmental impact. The study could generate further legal challenges when it is released.

Officials at NIAID admit that they tried to rush the project through after last autumn's US anthrax attacks. "So many things were happening so fast, there was a national emergency under way," says Jack Killen, NIAID's assistant director for biodefence research.



Not in my back yard: neighbours don't want a bioweapons lab added to the Rocky Mountain lab.

NIAID officials are hoping that the Rocky Mountain episode will improve their relations with the community. Staff at Rocky Mountain Laboratories have formed a committee to liaise with the community, and asked a member of Wulff's coalition to join it. But NIAID wants to build several new facilities, including two BSL-4 labs, and some bidders, such as the University of California, Davis, are situated close to large cities. Wulff says that her coalition may campaign for a more extensive environmental analysis of all US biosafety labs.

Bidders for the NIAID funding will be required to show that they have the support of the local community. NIAID officials hope that bidders can learn from more successful projects. The University of Texas Medical Branch at Galveston, for example, is building its own BSL-4 lab, for completion next year. It has gone ahead without significant resistance, partly because of an extensive education and outreach campaign by university officials. "The lesson is clear," says Killen. "You have to develop these facilities in collaboration with communities." ■

Traditional owners 'should be paid'

Quirin Schiermeier

Traditional knowledge can provide cheap leads for pharmaceutical companies looking for new drugs. Now a model law, unveiled at a conference last month on the South Pacific island of New Caledonia, could help to ensure that the indigenous communities involved are properly rewarded.

The law requires any company seeking to exploit traditional knowledge and culture to obtain the permission of the group that first developed it. Applications would be made through regional cultural authorities, which would be in charge of identifying and alerting the traditional owners. The owners would then have the right to reject the application or to negotiate an authorization agreement.

"It is a matter of self-respect for indigenous communities that their ownership of indigenous knowledge and culture is acknowledged in a fair way," says Kamal Puri, a lawyer at the University of Queensland in Brisbane, Australia, who co-authored the law.

The scheme is a response to cases in which drugs and cosmetic products have been derived from traditionally used plants without benefit to the indigenous community.

For example, smokebush (*Conospermum* spp.) has been used to develop an anti-AIDS drug by the Australian pharmaceutical company Amrad, based in Richmond, Victoria. Amrad obtained licences from the government of Western Australia and the US National Cancer Institute, where the plant's anti-HIV activity was identified, but neither contract requires payment to the aboriginal Nyoongah people, who use smokebush for medicinal purposes. In contrast, the government of Western Australia is likely to receive royalties in excess of \$100 million per year, says Puri, if Amrad successfully develops the drug.

"The Nyoongahs have used smokebush from time immemorial. Regrettably, no one has thought of providing any remuneration to those who first discovered the medicinal properties of the drug," says Puri.

The law was discussed by patent officials, lawyers and politicians from South Pacific states on 17 September. The government of Fiji has asked Puri for assistance in passing the law. Papua New Guinea, Samoa and Tonga are considering following suit, although some small island states may lack the legal set-up to enact the law. ■

International Space Station faces Russian funding shortfall

Washington NASA received a cry of distress last week. A top official in the Russian space industry warned that his country is close to defaulting on a key contribution to the International Space Station.

Moscow-based company RSC Energiya is due to build the Soyuz vehicles that are needed to ferry crew to the station, and the Progress vehicles used to take supplies there. But Valery Ryumin, a former cosmonaut and head of the station programme at RSC Energiya, said that his financially strapped company has not received the government money it needs to build the vehicles. Soyuz craft will also serve as the station's lifeboats, and NASA currently has no other means of evacuating the crew in an emergency.

In a letter to NASA space-station manager William Gerstenmaier, Ryumin proposed that the station be temporarily shut down. US officials played down any anxieties that Ryumin's letter may have caused. NASA says that it had "preliminary indications" of trouble with the station project, however, but would not respond until it received an official communication from Rosaviakosmos, the Russian space agency.



Sound waves: high-intensity sonar, used to track submarines, damages the ears of beaked whales.

Military manoeuvres leave whales high and dry

Madrid Spanish researchers investigating a mass stranding of whales in the Canary Islands late last month have found evidence that it was caused by a NATO military exercise.

The bodies of at least nine beaked whales were found on the beaches of Fuerteventura and Lanzarote on 24 September, while an exercise involving dozens of ships and submarines was taking place in the region. High-intensity sonar systems used to locate submarines have previously been shown to beach whales by damaging the animals' ears and disorientating them (see *Nature* 415, 106; 2002).

A research team at the University of Las Palmas, Gran Canaria, now says it has found injuries consistent with sonar damage in the

ear cavities and lungs of the dead whales. Michel Andre, a bioacoustics expert on the team, says the injuries are "strong evidence" that the military exercise was responsible, but he adds that further tests are needed to confirm this. The researchers hope to submit their full findings to the Spanish government by the end of this week.

Korea tightens rules after cloning claim

Seoul Claims that a South Korean woman has been made pregnant with a cloned human embryo have led the country's government to bring forward plans for stricter cloning legislation.

A new bill, known as the Life Ethics and Safety Measures, is currently being considered by the country's national assembly and is likely to be approved this autumn. The bill outlaws human reproductive cloning, but allows scientists wanting to create cloned embryos for research purposes to do so with the permission of a National Ethics Committee. It also allows embryonic stem cells to be extracted from embryos left over from *in vitro* fertilization, with the donor's permission, but these embryos will be required to be frozen for at least five years before use.

The move comes after the South Korean biomedical company BioFusion Tech claimed in July that it had implanted a cloned human embryo into a woman, and that she was then two months pregnant. BioFusion is an affiliate of Clonaid, a US company that promotes human cloning.

Camel joins top division of most endangered animals

Munich A wild Asian camel that is renowned for its extraordinary survival skills has been placed on the United Nations' (UN) list of most threatened species.

Numbers of the bactrian camel (*Camelus bactrianus servus*, which roams desert areas in China and Mongolia, are down to around 1,000, conservation experts told a meeting of

the UN Convention on Migratory Species last week in Bonn, Germany. The UN placed the species on the convention's Appendix 1 list of most endangered animals, a move that should help campaigners to raise money to protect the animals. Mongolian delegates pledged to combat camel hunting and illegal gold and oil mining, which are thought to have caused the decline in camel numbers.

"If nothing happens, the camels will become extinct within the next 25 years," says John Hare, an adviser to the convention and founder of the UK-based Wild Camel Foundation.

Other species newly listed in Appendix 1 include the great white shark (*Carcharodon carcharias*) and three species of whale.

PhD publication is key to what a Greek earns

Athens A contentious performance-related pay scheme for PhD students has been unveiled by the Greek Ministry of Education.

Under the plan, which is the first ever government funding for three-year university PhD placements, supervisors will be able to apply for a maximum of 80,000 euros (US\$79,000) for their students. Three-quarters of this is earmarked for publications — 7,000 euros can be earned for each of up to three publications in a peer-reviewed

journal or conference proceedings during the studentship. Two submitted, but not yet accepted, papers will be rewarded with 6,000 euros provided that the PhD is awarded. A maximum of only 10,000 euros is available for research materials.

The scheme, which makes use of European Union funding, has angered some Greek academics, who say it threatens the quality of basic research by making it a hostage to financial bargaining. They also complain that the ministry has not defined exactly what it means by a peer-reviewed publication.

Jan Hendrik Schön

Following the conclusion of an inquiry commissioned by Bell Labs (see Opinion, page 417, and News, page 419), readers should be aware that the following *Nature* papers co-authored by Jan Hendrik Schön have been shown to include falsified data. *Nature* is in touch with the authors to arrange for formal corrections to the scientific record.

Editor, *Nature*

Schön, J. H., Kloc, Ch. & Batlogg, B. Superconductivity at 52K in hole-doped C_{60} . *Nature* **408**, 549–552 (2000).

Schön, J. H. *et al.* Gate-induced superconductivity in a solution-processed organic polymer film. *Nature* **410**, 189–192 (2001).

Schön, J. H., Meng, H. & Bao, Z. Self-assembled monolayer organic field-effect transistors. *Nature* **413**, 713–716 (2001).

Schön, J. H. *et al.* Superconductivity in single crystals of the fullerene C_{70} . *Nature* **413**, 831–833 (2001).

Schön, J. H. *et al.* Superconductivity in $CaCuO_2$ as a result of field-effect doping. *Nature* **414**, 434–436 (2001).



Uncertain future: hunting has put the bactrian camel at risk of extinction within 25 years.

What difference does a genome make?

The malaria parasite's genome should provide a wealth of new scientific opportunities. But this may heighten tensions over how best to spend the scant resources allocated to malaria research and control. Declan Butler reports.

Holding out for a cure: how long will it be before increased knowledge of the malaria parasite translates into relief for those battling the disease?

The yellow international vaccination card tucked inside my passport describes malaria as “a serious and sometimes fatal disease, endemic in many tropical and subtropical countries. You cannot be vaccinated against it, but you can protect against mosquito bites, and take antimalaria tablets regularly. If you get a fever within two years after your return, tell your doctor you have recently been in a malarious country.”

Behind this bland warning lies a human catastrophe. Malaria kills more than a million people a year — or at least one person every 30 seconds — almost all of them in sub-Saharan Africa. Up to 500 million people suffer from the disease, with varying degrees of severity. Many are too poor to have access to protection and treatment, and existing drugs are rapidly losing their potency as the parasite evolves resistance to our limited pharmaceutical armoury.

In this week's *Nature*, an international team describes the complete genome sequence of *Plasmodium falciparum*, the deadliest malaria parasite¹. But how will this milestone influence the battle to defeat this devastating disease? Certainly, it will rejuvenate our understanding of the parasite's biology and its interactions with its human host. It will also generate new approaches and targets for drug and vaccine discovery.

But unfortunately, the availability of the *P. falciparum* genome does not herald the parasite's impending doom. The biggest bottlenecks often lie downstream, in getting funds and industrial expertise to move promising drugs and vaccines from the lab to the market. And, as experience with AIDS has shown, economic forces and inadequate health infrastructure can prevent Africa's poorest from benefiting from effective new treatments.

At least the outlook has improved in the past few years, with a series of new initiatives being launched to tackle malaria. In 1997, research agencies, charities and aid donors met with scientists in Dakar, Senegal, to explore ways forward. This led to the creation of the international Multilateral Initiative on Malaria, which aims to train African scientists and bridge the divides between bench researchers and field workers, and between developed and developing-world scientists. In 1998, Gro Harlem Brundtland, director-general of the World Health Organization (WHO), added political momentum by launching Roll Back Malaria, a global push in research and control that aims to halve deaths from the disease by 2010. And in July last year, the G8 leading industrialized nations pledged US\$1.3 billion to the Global Fund to Fight AIDS, Tuberculosis and Malaria.

But budgets for malaria research and control still fall massively short of what is required, making Brundtland's goal of rolling back malaria a distant dream. The world currently spends around \$200 million each year

on the disease, a figure that should be boosted this year by a further \$100 million for control from the G8-backed global fund. But this is a mere fraction of the sums made available for bioterrorism research in the wake of last year's anthrax attacks, and will do little to lessen the human misery caused by malaria.

Against this cash-strapped background, the flood of information that will flow from the *P. falciparum* genome is likely to exacerbate tensions between those who want to invest in developing new drugs and vaccines, and others who believe that the money would be better spent on improving existing countermeasures, such as insecticide-impregnated bednets. "Many would argue that children in Africa are dying for want of access to simple existing methods," says David Heymann, the WHO's executive director for communicable diseases. "Others would argue that more effective tools will serve the poor better."

Time trials

It's a delicate balance to strike, and one that must take into account the timescale over which information from genomics and other high-throughput biological analyses are likely to yield practical tools for malaria control. The *P. falciparum* genome should rapidly start churning out new drug targets that may move speedily into clinical trials. But old hands of malaria research, such as Louis Miller of the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, feel that it will take longer — perhaps two decades — to convert genomic information into effective vaccines. More ambitious schemes, such as releasing transgenic mosquitoes that are unable to transmit malaria (see page 429), lie further away still.

New targets for antimalarial drugs are badly needed. For decades, doctors in developing countries have depended on chloroquine or a combination of two drugs known as Fansidar to treat malaria. Both chloroquine and Fansidar are reasonably cheap, and were very effective. But they are now useless across large swathes of Africa because the parasite has grown resistant to them.

Of the 1,223 new drugs registered between 1975 and 1996, only three were antimalarials². And the handful of new antimalarial drugs currently being developed come from just three families of compounds — quinolines, antifolates and artemesinins — the first two of which are already succumbing to resistance.

The *P. falciparum* genome should yield new drug families. Putative enzymes, which are generally the most effective targets for drug development, can be identified from sequence data. And because the *P. falciparum* genome project has made its data freely available on the Internet, the opportunities can already be glimpsed. In 1999, German researchers identified an enzyme in a biochemical pathway for fatty-acid synthesis in

the parasite³, for which an inhibitory drug, fosmidomycin, had already been developed for an entirely different purpose — treating recurrent urinary infections. It is now in late-stage clinical trials as an antimalarial.

The outlook for vaccines is less certain, and acute funding shortages leave researchers divided over the importance of genomics in identifying parasite molecular signatures, or antigens, that the human immune system can be made to attack. Some researchers complain that they can't test all the antigens that have already been identified. "If there were an extra \$100 million to spend on malaria-vaccine research, I would allocate very little of it to exploring the parasite genome," says Adrian Hill of the University of Oxford, UK. He argues that the main problem is finding ways to make the human immune system respond more strongly to existing *P. falciparum* antigens. Another researcher puts it more bluntly: "If more money were spent on developing existing vaccine candidates rather than hyping the genome, I think lives would be saved sooner."

But Henk Stunnenberg of the University of Nijmegen in the Netherlands disagrees. He argues that vaccine developers need to be more ruthless in weeding out candidate vaccines at earlier stages, while exploring the wider range of antigens that the parasite's genome will provide. "If one is honest, a substantial number of promising candidates fell through in preclinical and clinical trials, but are still sold as steps forward," Stunnenberg claims. "Having more targets raises justified hope that we will succeed in developing effective vaccines."

Others argue that the genome will allow 'rational' vaccine design based on antigens' function and the way in which they stimulate the immune system. But Louis Schofield, a malaria researcher at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, warns against excessive optimism. "A plethora of new vaccine candidates will be useless without a thorough understanding of total parasite biology," he says. That will mean exploring both the *P. falciparum* and human genomes to gain insight into the interaction between host and parasite. It may also require require malaria research labs to retool and restaff to emphasize bioinformatics, proteomics and large-scale studies of protein-protein interactions — not least because *P. falciparum* undergoes profound changes in patterns of protein production at different stages of its life cycle.

While bench researchers debate how the *P. falciparum* genome should influence their priorities, other initiatives are trying to address the bottleneck that lies downstream: the difficulty of pushing to market promising drug and vaccine candidates that would be seen as commercial liabilities under the pharmaceutical industry's usual rules — where profits come from drugs that can be

mass-marketed at developed-world prices.

The Geneva-based Medicines for Malaria Venture (MMV), for instance, was launched in 1998 to kickstart discovery and development of new families of antimalarial drugs. Its goal is to raise \$30 million annually from public and private donors to register a new, affordable antimalarial every five years. Drug companies have provided project-management expertise and technologies such as combinatorial chemistry and high-throughput screening. With seven drug-discovery and five development projects under way, the MMV now has the largest antimalarial drug pipeline since the Second World War, says its chief executive officer, Christopher Hentschel. So far, however, the scheme has attracted funds of just \$15 million per year — a fraction of what is needed — with \$5 million coming from one individual, Microsoft chairman Bill Gates.

Gates' philanthropy is also underwriting the Malaria Vaccine Initiative, based in Rockville, Maryland, which aims to do for vaccines what the MMV is doing for drugs. It was launched in 1999 and is supported by \$50 million from the Bill and Melinda Gates Foundation.

Setting priorities

Unfortunately, drug- and vaccine-development efforts must compete for limited funds with fundamental research into the malaria parasite and the costs of programmes to control malaria in the field. And given the enormity of the current death toll, many public-health experts argue that the latter must be given high priority (see Correspondence, page 431).

Tikki Pang, the WHO's director of research policy and cooperation, argues that this contentious debate must be informed by something that is currently lacking: rigorous, independent cost-benefit analyses of the value of expanding and improving existing countermeasures, such as bednets and education, compared with investing for the future in genomics-based research to produce new drugs and vaccines.

Almost everyone can agree, however, that it would be desirable to help the countries most severely afflicted by malaria to commit the resources needed to combat the disease, and to train their researchers to get involved. Sadly, much remains to be done. The Multilateral Initiative on Malaria has made progress, but many African researchers argue that other developed-world research groups and funders need to rethink their priorities. Kevin Marsh, director of the Kenya Medical Research Institute-Wellcome Trust unit in Kilifi, on the Kenyan coast, argues that donors have "massively underestimated" the needs, and too often provide short-term funding with little strategic vision. This has created "a group of high-level research assistants, but not international-calibre scientists", he says.

Genomics is sometimes portrayed as a

great leveller, allowing any scientist with Internet access to query sequence databases and make discoveries. But African scientists complain that so far they have been sidelined. "First-world scientists pursuing genomic research are generally not working in collaboration with endemic-area scientists and local institutions," agrees Gerald Keusch, director of the US National Institutes of Health's Fogarty International Center in Bethesda, Maryland. "If the promise of the parasite genomes is to be realized, there needs to be a greater effort to develop trusting, transparent and collaborative relationships with institutions in the countries where the fruits of malaria and mosquito genome research will be used."

Inclusive effort?

Winston Hide, director of the South African National Bioinformatics Institute in Cape Town, is annoyed that efforts to sequence the genomes of *P. falciparum* and its mosquito vector, *Anopheles gambiae*⁴, did not include developing-country groups. Although Hide accepts that most African labs are not equipped to undertake the sequencing itself, he says they could have been involved in the analysis, which would also have given them early access to unpublished information.

Malcolm Gardner, lead author of the *P. falciparum* genome paper, says that the consortium has gone out of its way to put preliminary sequence data on the Internet. The *Plasmodium* Genome Database has also been made available on CD-ROM to take into account poor Internet access in developing countries, he adds.

But there are precedents for using genome projects to build scientific infrastructure in a developing country — albeit one richer than those in sub-Saharan Africa. In July 2000, some 30 labs in the Brazilian state of São Paulo published the first complete genome of a plant pathogen, the bacterium *Xylella fastidiosa*, which causes disease in orange trees⁵. And with the genomes of *P. falciparum* and other tropical parasites now becoming available, the South African government has pledged US\$40 million over ten years to create a national bioinformatics network centred on Hide's institute, which he hopes will become a hub for the whole of Africa.

Disappointment over the involvement of African scientists in malaria research is nothing, however, compared with the despair with which public-health experts view the disparity between the demand for malaria control in Africa and the supply of resources and trained staff. The WHO Commission on Macroeconomics and Health, an expert panel set up by Brundtland that recently reported on global health needs⁶, argued that, to control malaria effectively, funding for its control needs to rise to \$2.5 billion annually by 2007, and to \$3.1 billion by 2015. At present, only a handful of sub-Saharan African countries have been able to raise their spending on malaria control to

the level needed even to meet the more modest goals of Roll Back Malaria.

"Control will not make much progress unless the international community is willing to tackle the gross underfunding and major problems of health-service infrastructure in Africa," concludes Anne Mills, a health economist at the London School of Hygiene and Tropical Medicine and a member of the WHO Commission on Macroeconomics and Health. "Individual disease initiatives can only go so far when basic health infrastructure is weak or non-existent."

Mills' message applies across the entire spectrum of malaria research and control. While biomedical scientists and public-health experts argue over the relative merits of investing in genomics, drug and vaccine development, or the application of existing control measures, what's really needed is more money across the board.

With the resources currently available, say health economists, talking about defeating malaria is like promising to build a \$100-million skyscraper with just \$1 million in the bank. The stark truth is: if we don't bankroll the effort, we won't roll back malaria. ■

Declan Butler is Nature's European correspondent.

1. Gardner, M. J. et al. *Nature* **419**, 498–511 (2002).
2. Trouiller P. & Olliaro P. L. *Int. J. Infect. Dis.* **3**, 61–63 (1999).
3. Jomaa, H. et al. *Science* **285**, 1573–1576 (1999).
4. Holt, R. A. et al. *Science* **298**, 129–149 (2002).
5. Simpson, A. J. G. et al. *Nature* **406**, 151–157 (2000).
6. *Nature* **415**, 1 (2002).

Multilateral Initiative on Malaria

♦ <http://mim.nih.gov>

Roll Back Malaria

♦ www.rbm.who.int

Global Fund to Fight AIDS, Tuberculosis and Malaria

♦ www.globalfundatm.org

Medicines for Malaria Venture

♦ www.mmv.org

Malaria Vaccine Initiative

♦ www.malariavaccine.org



Despite the use of insecticidal bednets (top), drug programmes (middle) and pesticides (above), many Africans remain at risk from malaria.



Mosquitoes minus malaria

If wild populations of the mosquito that transmits malaria were replaced with insects rendered harmless by genetic engineering, the disease could finally be defeated. But that remains a big 'if', as Tom Clarke finds out.



S. STAMMERS/SPL

For millions of people in sub-Saharan Africa, the whine of *Anopheles gambiae* is the music of death. This mosquito spreads the malaria parasite *Plasmodium falciparum*, but if it could be genetically altered to block the cycle of transmission, it would hold no fear. Its bites would be just itchy irritations, its sound merely a mild annoyance.

Following recent advances, genetic engineers are now confident that they can make *A. gambiae* malaria-proof, or 'refractory'. And the publication in *Science* this week of the insect's genome sequence¹ is bound to heighten interest in tinkering with its genetic make-up. But even if a transgenic refractory mosquito can be produced, how can we ensure that it completely infiltrates wild populations? Would the genetic modification remain stable? And what are the public-health implications if the scheme is only a partial success?

Answering these questions will take years of meticulous lab and field research, which at any turn could throw the ambitious scheme off track. Even scientists who support the effort accept that it has to be viewed as a long shot. "I'm not quite ready to hitch my cart to the transgenic wagon," says John Edman, director of the Center for Vector-Borne Diseases at the University of California, Davis.

Some researchers argue that money would be better spent on other approaches to malaria control. But for others, the potential reward remains too great to ignore. Three times in the past two years, most recently in Wageningen, the Netherlands, in June, geneticists, medical entomologists and public-health experts have gathered to discuss the research that would be required if we are to create a malaria-free future through the release of transgenic refractory *A. gambiae*.

Just weeks before the Wageningen meeting, a paper published in *Nature*² illustrated the potential. Researchers led by Marcelo Jacobs-Lorena of Case Western Reserve University in Cleveland, Ohio, showed that

adding a synthetic gene for a peptide called SM1 into the related species *A. stephensi* almost completely disrupted the mosquito's ability to transmit *Plasmodium berghei*, which causes malaria in mice. The peptide resembles a receptor in the mosquito's gut and salivary glands that is thought to allow the parasite to burrow from one to the other. Flooding the mosquito's body with a similar molecule seems to confuse the parasite, preventing its progress.

Meanwhile, Andrea Crisanti and his colleagues at Imperial College, London, are working on genetic manipulations that direct *A. gambiae*'s immune system against *P. falciparum*. "Once the best target is identified it would be a matter of months before we have a modified mosquito," Crisanti asserts. Other, more generic approaches are also being considered, such as genetic modifications that render the offspring of the engineered insects sterile.

Jump to it

But creating genetically modified mosquitoes in the lab is just the start. How can they be made to spread their altered genes through wild *A. gambiae* populations? It's a tall order, but genetic elements that could fit the bill exist naturally in insect populations. These transposons, or 'jumping genes', copy themselves and move around the genome, being passed down the generations. Once they enter a population, they can spread rapidly through it.

There are many examples, but the most famous is the P-element, discovered in the fruitfly *Drosophila melanogaster* in the late 1970s. Evidence suggests that the P-element jumped into *D. melanogaster* from a close relative less than 100 years ago. It is now found in all wild populations of the fly worldwide. So the idea is to slot genes for refractoriness into a suitable transposon, insert this package into a mosquito and then watch it spread.

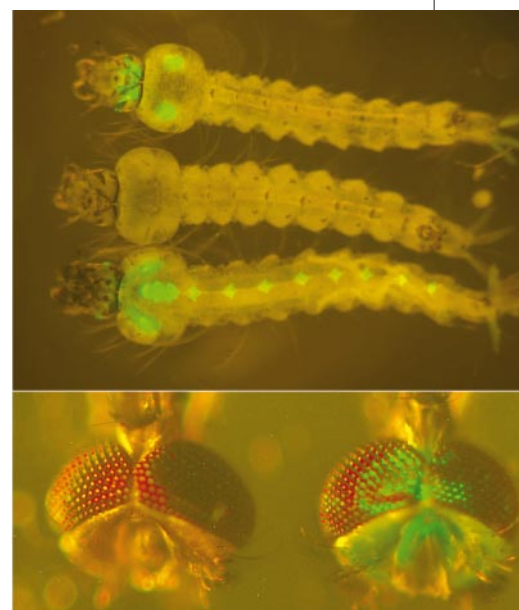
Here, however, genetic engineers' ambition may be thwarted by natural selection.

Wild *A. gambiae* have been adapting to their environment for many millions of years. But breed them in the lab for several generations and their evolutionary fitness may plummet. Their ability to compete for food, find mates and avoid predators could all be compromised.

Even if lab-reared strains are continually crossed with vigorous wild-type mosquitoes, the genetic modification itself may also reduce their fitness, as unpublished preliminary work indicates. "There is a cost to mosquitoes of carrying a foreign gene," suggests Peter Billingsley, a molecular physiologist who studies malaria transmission at the University of Aberdeen, UK.

On the positive side, the introduced genes would free the transgenic mosquitoes from the burden of carrying *P. falciparum*. There is some evidence that the parasite has a

REF. 2



Shining examples: a transgene makes these mosquito larvae (top) and adults glow green.

▶ detrimental effect on *A. gambiae*'s fitness^{3,4}, but scientists would want to conduct extensive experiments to see whether this would outweigh any fitness deficits before going ahead with field releases.

Population geneticists also warn that the refractoriness genes could become unhitched from their transposon and be left behind. "Once relieved of its burden, the transposon will steam on ahead," says Chris Curtis, a medical entomologist at the London School of Hygiene and Tropical Medicine. And after a particular transposon has spread through a population, it can't be used again.

Perhaps the biggest worry, however, is whether the modification would remain stable in the long term. It could sweep through the global population only to be selected out after a number of years. Alternatively, the malaria parasite might develop resistance to the modification — just as it has to anti-malarial drugs. Both scenarios have profound implications for public health. A temporary halt in malaria transmission could result in people losing all natural immunity, rendering the disease even more devastating once the parasite returns. "If malaria were eliminated from Africa for a short period of time, the consequences could be terrible," says Andrew Spielman, a medical entomologist at Harvard University.

Potential pitfalls

Other concerns are more speculative. As a transposon jumps around the *A. gambiae* genome, for example, it could disrupt other genes, with unpredictable results. It's extremely unlikely, but what if a transposon gave mosquitoes the ability to transmit other diseases? Changing the relationship between mosquito and malaria parasite also could have unforeseen ecological and evolutionary impacts. Could the parasite adapt quickly enough to use another vector to continue its deadly life cycle? If sterility genes were used to eliminate *A. gambiae*, would other species, which only occasionally bite humans, move in and replace them? Could these new mosquitoes transmit malaria?

Given the catalogue of unknowns, and the potential consequences if the plan were to go awry, experts say that an intensive, multi-stage research programme is needed, similar to the clinical trials of safety and efficacy carried out for experimental drugs. Laboratory studies would first assess the stability, fitness constraints and safety of various genetic modifications. Harmless and well-studied transgenes, such as that for jellyfish green fluorescent protein, could be used at first.

Later on, experiments would move on to the genes that would be used for real, in a controlled environment that mimics the wild as closely as possible. A gigantic green-



Golden opportunity: Hawaii and its introduced mosquito *Culex quinquefasciatus* (inset) could provide an ideal test for the transgenic strategy.

house in Kenya, dubbed the 'Malariasphere', could provide a suitable testing ground. Operated by the International Centre of Insect Physiology and Ecology in Nairobi, the Malariasphere contains mock-ups of huts, forest and breeding puddles for mosquitoes. It is currently being used to study the dynamics of *A. gambiae* populations and their ability to spread malaria.

While such basic research is under way, researchers would need to investigate suitable sites for transgenic releases. Extensive baseline information on local populations of mosquitoes and the burden of malaria in the area would be needed for field workers to be able to tell if a later release of refractory

The biggest worry is whether the modification would be stable in the long term.

mosquitoes is having the desired effect.

Before giving the eventual go-ahead, however, authorities may want to see the principle demonstrated in the field under circumstances in which the stakes are not so high. Some argue that it would be best to begin on an island with no malaria, working with a mosquito that does not transmit the disease or bite humans. This way, the dynamics of an artificial gene moving through a population could be studied more safely.

Natural setting

One intriguing idea, championed by Edman, would be to use the geographically isolated Hawaiian islands as a natural laboratory. They are host to the mosquito *Culex quinquefasciatus*, which transmits a form of malaria to birds. The mosquito was accidentally introduced in the 1820s and is currently accelerating the decline of the islands' native birds. Releasing transgenic refractory *C. quinquefasciatus* could help to test many of the ecological unknowns while taking

advantage of the excellent scientific infrastructure in Hawaii — which does not exist in many regions afflicted by *P. falciparum*. And if successful, the experiment would help solve a genuine problem.

Conducting the research needed to test the idea of transgenic refractory mosquitoes won't be cheap. But Kate Ault-

man, an official at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, who is responsible for funding projects on malaria, says that if a reasonable research agenda were constructed, money could be found. "I look forward to getting many proposals," she says.

In any case, say many medical entomologists, the wealth of information on the ecological dynamics of *A. gambiae* populations that would be gained from the research would be valuable in informing other efforts to combat malaria by attacking its vector. "Many of these issues should be high on the agenda," argues Tom Scott, a medical entomologist at the University of California, Davis, "whether we go for an eventual release or not." ■

Tom Clarke works in Nature's news syndication team.

1. Holt, R. A. *et al. Science* **298**, 129–149 (2002).
2. Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. & Jacobs-Lorena, M. *Nature* **417**, 452–455 (2002).
3. Anderson, R. A., Knols, B. J. G. & Koella, J. C. *Parasitology* **120**, 329–334 (1999).
4. Ahmed, A. M. *et al. Oikos* **97**, 371–377 (2002).

Integrated programme is key to malaria control

Genome publication must not divert attention from basic science and public-health goals.

Sir — After decades of relative neglect, malaria control is again high on the international research and public-health agenda. The exciting simultaneous publication of the sequence and functional analysis of the genomes of two malaria parasites, *Plasmodium falciparum* and *P. yoelii*, in this issue of *Nature*¹, and mosquito vector *Anopheles gambiae* in *Science*² should facilitate accelerated development of new drugs, insecticides, vaccine candidates and engineering of malaria-resistant mosquitoes.

The future availability of such powerful tools is a captivating prospect, but their allure must not divert attention from potential basic science and public-health goals that remain grossly underdeveloped. Our understanding of malaria-transmission ecology lags well behind our understanding of climate change or biodiversity loss. Most worrying of all is the absence of proven environmental interventions where they are most needed, in particular in sub-Saharan Africa.

Although the effectiveness of preventing mosquito proliferation in Asia, Europe and the Americas has been well known for 80 years (see, for example, ref. 3), spectacular historical successes against tropical African vectors should not be ignored. Successful programmes to control malaria transmission by *A. gambiae* — the predominant vector in sub-Saharan Africa — were initiated in the 1930s and 1940s and their effects sustained for two decades or more in Brazil, Egypt and Zambia^{4,5}. There were three common features to these

outstanding successes: they had integrated control, emphasizing environmental management and regular insecticidal suppression of larval stages of the vector; they used rigorous surveillance and adaptive tuning of the intervention package over time; and they employed multisectoral programme staff with expertise in clinical, ecological, entomological and epidemiological aspects of malaria, and in land and water management. The challenge now is sub-Saharan Africa itself, where the frequency of infectious bites means that vector control has not so far worked.

We are concerned that the current global malaria control campaigns (see, for example, ref. 6 for an overview), and in particular the emphasis on genomics and vaccine development likely to be stimulated by the new results, may delay the re-adoption of integrated methods, as happened with the euphoria that greeted the introduction of DDT as a control method decades ago. In addition to laboratory-derived tools, integrated malaria control in Africa will rely on basic and applied environmental science *in situ*.

None of these skills or tools can be applied without the essential infrastructure embodied by the second and third criteria outlined above. This is where contemporary public-private partnerships could make a profound difference in supporting and facilitating integrated malaria control, as in the successful programmes of the past. The Global Health Initiative of the World Economic

Forum is playing an important role in fostering such partnerships⁷. The corporate sector — for example, Exxon-Mobil in Cameroon and Chad; British Petroleum in Angola; and Konkola copper mine in Zambia — is implementing integrated malaria control today.

Consequently, a detailed business plan is required for malaria control on a broader scale, analogous to that produced earlier this year by WHO for tuberculosis. The strengthening of human resources and infrastructure in Africa, together with the increased public funding available for malaria research and control initiatives, will enable successful re-adoption of proven approaches from the past while we eagerly await the benefits of the genomics revolution.

Jürg Utzinger*, **Marcel Tanner†**,
Daniel M. Kammen‡, **Gerry F. Killeen†**,
Burton H. Singer*

*Princeton University, Office of Population Research, Princeton, New Jersey 08544, USA

†Swiss Tropical Institute, PO Box, CH-4002 Basel, Switzerland

‡Renewable and Appropriate Energy Laboratory, Energy and Resources Group, University of California, Berkeley, California 94720, USA

1. *Nature* **419**, 498–542 (2002).
2. Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
3. Kitron, U. & Spielman, A. *Rev. Infect. Dis.* **11**, 391–406 (1989).
4. Killeen, G. F., Fillinger, U., Kiche, I., Gouagna, L. C. & Knols, B. G. J. *Lancet Infect. Dis.* **2**, 618–627 (2002).
5. Utzinger, J., Tozan, Y., Doumani, F. & Singer, B. H. *Trop. Med. Int. Health* **7**, 657–677 (2002).
6. *Insight: Malaria Nature* **415**, 669–715 (2002).
7. World Economic Forum, Global Health Initiative: Resource Paper (World Economic Forum, Geneva, 2002). www.weforum.org/pdf/Initiatives/GHI_Resource_Paper.pdf

How serendipity led to an early treatment

Sir — In her review of Mark Honigsbaum's book *The Fever Trail* (*Nature* **418**, 820–821; 2002), Sandra Knapp asks how the *Cinchona* tree's properties came to be discovered, given that malaria did not exist in the New World until the Europeans arrived there. The book does not answer this question.

Several species of the *Cinchona* tree (which was called “quina-quina” by the Indians) grow on the warm, moist slopes of the Andes above 1,500 metres, where the *Anopheles* mosquito does not survive. The quina-quina trees were thought by some people to be poisonous, owing to the bitter taste of the bark. Yet plants that are unpleasant to the taste or even harmful if taken unwisely may be used, in the right

form and dosage, as medicines.

Therefore it is plausible to suppose that the bark of quina-quina trees has also been used therapeutically by South American Indians since pre-Columbian times for chills and fevers. (One legend tells of an Indian who, burning with fever, drank from a jungle pool despite the bitter taste that revealed it was contaminated by the neighbouring quina-quina trees, and whose fever then abated.) The Jesuits probably learned about the anti-fever properties of this bark from the Indians and tried it, successfully, to treat malaria — hence the remedy's popular name of ‘Jesuit's bark’.

According to a story that is widely accepted in Western countries, the use of the bark for malaria is a case of serendipity (see R. M. Roberts, *Serendipity: Accidental Discoveries in Science*, 6–10; Wiley, New York, 1989). The Countess of Chinchon

(1576–1639), wife of the Viceroy of Peru, was said to have been cured of malaria by taking an extract from the bark of a Peruvian tree. She was also said to have carried some bark back with her to Spain in 1638, thereby introducing its use in Europe. This part of the story, at least, is false, because the Countess died in Cartagena de Indias, Colombia, without ever returning to Spain — though she was immortalized by Linnaeus when he gave the tree its botanical name.

In 1633, the monk Calancha, who accompanied the Spanish conquistadors, introduced the use of the remedy *Cinchona* to Europe (for more information, see R. B. Silverman, *The Organic Chemistry of Drug Design and Drug Action*, 1–2; Academic, San Diego, 1992).

Pedro Cintas

Department of Organic Chemistry, Faculty of Sciences-UEX, E-06071 Badajoz, Spain

Better safe than sorry

Is the precautionary principle a useful guide to action?

The Precautionary Principle in the 20th Century: Late Lessons from Early Warnings

edited by Poul Harremoës, David Gee, Malcolm MacGarvin, Andy Stirling, Jane Keys, Brian Wynne & Sofia Guedes Vaz
Earthscan: 2002. 284 pp. £48, \$79.95 (hbk); £17.95, \$29.95 (pbk)

Roger Pielke Jr

When President George W. Bush rejected US participation in the Kyoto Protocol, he stated: "We must be very careful not to take actions that could harm consumers." This statement reflects the application of the 'precautionary principle', defined by the editors of this book as "acting before there is strong proof of harm, particularly if the harm may be delayed and irreversible" — or, in the language of common sense, "better safe than sorry".

Bush's statement reflects his concern that policies focused on environmental protection may cause harm to the economy. By contrast, proponents of the Kyoto Protocol are concerned that policies focused on economic development may cause harm to the environment. This example illustrates the limited usefulness of the precautionary principle as a guide to action. Both Bush and his opponents can appropriately use 'better safe than sorry' to justify their perspectives on the Kyoto Protocol — it provides no mechanism to reconcile disputes about how to make trade-offs between competing values.

An alternative would be to specify in advance which 'harms' should trump others; for example, the economy should always win out against the environment, or vice versa. But when the precautionary principle is associated with only one set of valued outcomes, an inevitable outcome occurs,

which Tim O'Riordan characterizes in the foreword: "Precaution is a fully politicised phenomenon."

The book is organized around 14 case studies: fisheries, radiation, benzene, asbestos, polychlorinated biphenyls, halocarbons, diethylstilboestrol, antimicrobials, sulphur dioxide, methyl tertiary-butyl ether, chemical contamination of the Great Lakes, trybutilin compounds, hormones and mad cow disease. Each case study addresses four key questions. When was the first credible scientific early warning of potential harm? When and what were the main actions or inactions on risk reduction taken by regulatory authorities or others? What were the resulting costs and benefits of the actions or inactions, including their distribution between groups across time? And what lessons can be drawn from the affair that may help future decision-making?

This organization leads to an extended exercise in affirming the consequent. Each of the short case-study chapters covers a subject that, with the availability of hindsight, involves some human activity (primarily the production and use of chemicals) that has since been proven to harm the environment and/or people. A lesson repeated at the end of most of the case studies is that we should have known better and acted more promptly to regulate the harmful activity.

But this approach to the selection of cases overlooks two critically important outcomes. First, what about cases in which uncertainty existed about a particular activity for which so far no harm has been detected that outweighs the benefits? How might the approach recommended by the authors have been implemented in such cases? If one objective of the book is to suggest an alternative mecha-

nism of decision-making in the context of uncertain outcomes, it is a serious oversight to ignore cases that begin shrouded in uncertainty yet develop with benign outcomes.

By selecting only cases in which negative outcomes resulted, the authors imply either that uncertainty always results in harm (which it clearly does not), or that a precautionary approach makes sense only in those cases where harm eventually occurs. But estimating the potential for harm is the subject of quantitative risk assessment and prediction, the standard methods of business-as-usual that the authors claim to want to move beyond.

Second, what of the cases in which regulatory or other precautionary action was taken that turned out to be exceedingly costly or unnecessary? The editors rather unconvincingly dismiss the latter cases, claiming that despite their efforts, "no suitable examples emerged". This is a difficult claim to accept, as candidate examples of such cases are manifest. Examples found on the pages of newspapers recently include mammography, US federal flood insurance, oestrogen therapy, the medical use of marijuana and other illegal drugs, airport security screening and forest-management policies.

But even with the logical flaws and selective set of case studies, the 12 lessons recommended at the end of the book are eminently reasonable and do indeed suggest an alternative to business as usual. But there is no direct connection between the cases selected and the lessons. The reader might rightly wonder where these valuable lessons actually came from, and why the authors did not see fit to organize their work around such wisdom.

In trying to shoehorn the various cases into the ill-fitting precautionary principle,



On the front line: tightening airport security in the war against terror creates queues but has yet to provide a solid return on its considerable investment.

rather than around a diverse set of practical lessons borne from experience, the authors have limited this book's contribution to an alternative view of how decision-makers might confront complex and contested environmental and science policies. ■

Roger Pielke Jr is at the Center for Science and Technology Policy Research, University of Colorado/Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado 80309, USA.

Mind over matter?

The Ingredients: A Guided Tour of the Elements

by Philip Ball

Oxford University Press: 2002. 228 pp. £13.99

Gautam R. Desiraju

Another book on the chemical elements? It is daunting to write about such a familiar topic but Philip Ball has the credentials to do so. A consultant editor for *Nature*, he has written extensively about science in addition to broadcasting on television and radio.

The main hurdle to writing about the elements is that chemistry is all about molecules and not elements, just as a song is about tunes and not notes and a sentence is about words and not letters. There are those who feel that even molecules do not provide a full understanding of chemistry, and that molecular assemblies, rather than individual molecules, are the irreducible leitmotifs of function in both materials science and biology. In such a scenario, the 120 or so chemical elements belong more to physics than to chemistry and, in their austerity, remind us of a time when chemistry was closer to atomic physics than to functional biology.

So are the elements relevant today? Ball has deflected this question in two ways in this provocative book, which is surely worth a read by both general readers and chemists. He has been quite subjective in his choice of elements, and the book is clearly a 'guided' tour. Secondly, he has taken poetic licence in defining the term 'element'.

The first of these two approaches cannot be questioned. So much has been written about the elements that a new book need not be just descriptive or comprehensive. To a physicist, the formation of any of the known elements is merely a reaffirmation of the rules of quantum mechanics. Yet the distribution of the 92 elements found naturally on the surface of our planet is quite skewed. No more than 30 are of widespread occurrence in the mineral world. The entry of an element into the 'biological club' is even more stringent, being largely restricted to around six members of moderate temperament that have low atomic



Golden touch: the tale of King Midas highlights the allure of gold for human societies.

number and a prime location in the middle of the periodic table. These attributes lead readily to covalency, which in turn favours specificity and reversibility in the reactions in which the corresponding molecules can participate. Carbon, cerium and californium are not equal partners, and Ball has done well in not treating them in the same way.

The elements that the author has chosen to highlight are excellent selections. Gold and oxygen would have been my choices, too. Gold has always fascinated humans and will always continue to do so. Ball's account of this "most useless of metals" is arresting as it brings together alchemy, chemistry, physics, metallurgy, a little biology, history, geography, economics, sociology and culture, in what is truly the best chapter of the book. The author reiterates that it is metals, rather than any of the other elements, that have so definitively encouraged or limited the scope of human activity. Coinage metals fascinated man long ago, then in the twentieth century the focus shifted to metals such as chromium, tungsten and the platinum group. Today, geopolitical strategy is driven by the availability or accessibility of titanium, zirconium and thorium. Clearly, nature has bestowed her favours unequally in her distribution of the metallic elements, and we can expect that humans will continue to be as unprincipled, as ruthless and in the end as silly as they always have been in their quest to gain access and control of metals.

But if gold is the sovereign of the periodic table, oxygen is surely the prime minister. Ball's account of this element is mostly historical, however, and much of what he says about the scientific interactions or otherwise between Antoine Lavoisier, Joseph Priestley and Carl Scheele, for example, has been described elsewhere. The role of oxygen in maintaining the various biochemical cycles

is also well known. I would have liked to have seen more in this chapter. Oxygen is too pivotal an element to be treated lightly and needs to have been addressed from a modern viewpoint. It is the only element, apart from fluorine, that can react with so many others. Unlike fluorine, however, it is not so reactive as to become exotic, and it is more flexible than fluorine in the kinds of bonds it forms. Oxygen can, in effect, feel the pulse of the rest of the elements quite accurately, and it does this in increasingly subtle ways, such as by participating in hydrogen bonds.

The organization of the elements into the periodic table is treated competently, as is the synthesis of new elements in the 'atom factories' in Berkeley and Dubna, and these chapters contain a lot of information. Ball's description of isotopes is particularly good, and I do not believe that I have seen such an imaginative handling of this usually stodgy topic. His breezy journey through the landscape of some technologically important elements (iron, silicon, palladium, the lanthanides and the noble gases) provides a lively end to the guided tour.

It is in the second of his stratagems that Ball becomes more controversial. When he states that "the story of the elements is not simply a tale of a hundred or so different types of atoms... it is a story about our cultural interactions with the nature and composition of matter", he moves beyond what is strictly scientific towards a looser and softer mode of thought. The holistic approach of Aristotle led to the quartet of earth, fire, air and water. Add a fifth — ether — and we obtain the 'panchabhoota', or five elements, of the Hindu canon; the Absolute is worshipped in these manifestations to this day.

Ball refers to Plato, Galen, Leonardo, Shakespeare and T. S. Eliot to illustrate that these holistic concepts have satisfied humankind's needs over time. But it is an absolute fact that these interpretations are scientifically wrong, just as Dmitry Mendeleev, Lothar Meyer and the scientists who preceded them for a hundred or so years were scientifically correct about the nature of the elements. The periodic table is one of the grandest intellectual accomplishments of the scientific world, and in trying to make the story of the elements more inclusive by interpreting a scientific subject from the viewpoint of the humanities, Ball is treading on unexplored and possibly treacherous ground. Conversely, critics of physics and chemistry maintain that all of the important problems in these subjects are essentially solved. Is this book then an implicit recognition of this fact? ■

Gautam R. Desiraju is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

Stories of the Invisible: A Guided Tour of Molecules (Oxford University Press, £7.99), also by Philip Ball, is now published in paperback.

Shaking the tree of life

The Nature of Diversity: An Evolutionary Voyage of Discovery
by Daniel R. Brooks & Deborah A. McLennan

University of Chicago Press: 2002. 680 pp.
\$85, £59.50 (hbk); \$35, £24.50 (pbk)

Sean Nee

"The Nature of Diversity is both phylogenetics and the environment, inextricably interwoven, rich in anachronisms and serendipity." Anachronisms and serendipity? This weird summary of the book is on the back cover. In fact, by virtue of its weirdness, perhaps it is an excellent description of a book with content such as this: "Darwinism is summarized by the Jagger principle: You can't always get what you want, but if you try sometime, you just might find you get what you need." This *aperçu* is actually highlighted for our special attention.

Biological diversity can be described by phylogenies — family trees that show the relationships between species. Such fundamental evolutionary information can be used for many purposes, such as identifying biological or environmental variables that promote speciation. *The Nature of Diversity* aims to tell us how to construct phylogenies, and then the sorts of things that we can then do with them. Brooks and McLennan give us an idiosyncratic view of these subjects, to say the least.

To construct a phylogeny, you need data pertaining to the species' relationships and a technique to extract the phylogenetic information from that data. Readers of *Nature* will have encountered many papers that present phylogenetic trees constructed by maximum-likelihood methods and, more recently, by bayesian methods, applied to molecular sequence data. To get a good phylogeny quickly, all you need to do is feed the output from your automated sequencer into a standard phylogenetic computer package and presto! — you have a phylogeny constructed using some of the most powerful and advanced techniques, theoretical and empirical, of the past 20 years.

But if you wish to learn something about these modern techniques you must look elsewhere. All you will learn from this book about maximum-likelihood methods is that the authors do not like them much, whereas bayesian methods do not seem to exist at all. I did not carefully study the authors' preferred recipe for phylogeny construction; suffice it to say that it has such ingredients as "Hennig's auxiliary principle" and "Remane's criteria". *Nature* readers will rarely encounter such arcana.

The authors are not crazy either about molecular sequence data. This is unfortu-

nate, as molecules nullify some apparently grand concerns of the phylogenetic community, judging by the space given to their discussion. For example, what is a species? This question has different meanings in different contexts. Here's one. Consider a simple phylogeny consisting of, say, just humans, chimpanzees and the lineage that gave rise to us both. Certainly, humans and chimpanzees are different species. But what about the ancestral lineage? Is it a distinct, third species? It seems that there are different schools of thought on this question of questionable importance, differing in how they react to a situation in which this ancestor has no 'characters' that are different to those of, say, chimps. Well, even though we do not have the sequence data, we can know that the lineage had differences in molecular characters, so we need trouble ourselves no further and can turn to more interesting questions.

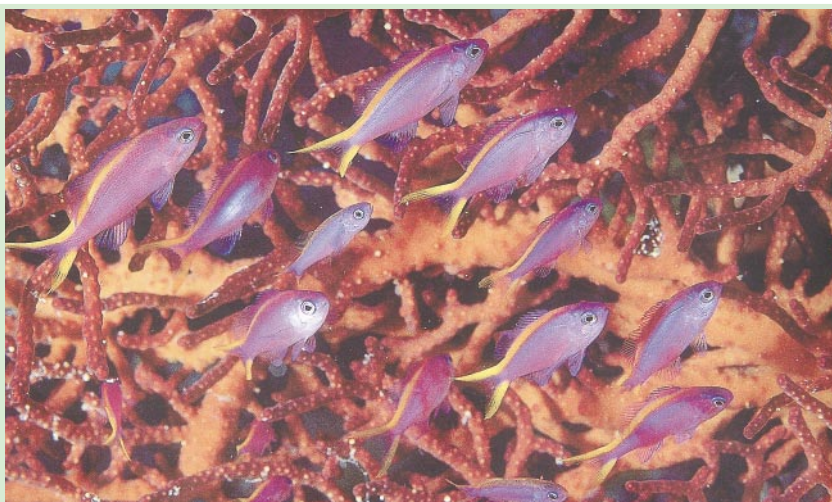
This book does become useful in the chapters that summarize numerous studies that have used phylogenies to make inferences in many different areas of enquiry, such as biogeography, coevolution and adaptive evolution. I learned of many interesting studies that I would probably not otherwise have encountered. But problems frequently arise in the synthesis of the material. Consider, for example, questions about 'radiations'. Say you have two taxonomic families of the same age, but one has twice as many species as the other. Is that remarkable, or could it reasonably have arisen by chance alone, even if they had the same speciation probabilities over time? Another question: I have numerous pairs of families, and in each of these pairs, one family has only blue species and the other only green

species and, in each pair, the blue family has more species than the green family. Is this remarkable, or might the apparent association between colour and species richness have readily arisen by chance alone? Alas, these two questions are inextricably conflated in this book, although they are actually completely different and must be kept well separated for correct analysis.

This big book has about 3,000 references, but this abundance is generated largely as a result of long lists of randomly important references to a topic, rather than just a few references to definitive works. So, for example, on ecological diversity indices we are not referred to Anne Magurran's excellent monograph *Ecological Diversity and its Measurement* (Princeton University Press, 1992). Instead we get a seemingly arbitrary list including such unicorns as a French paper from 1908 published in "*Bull. Soc. Vaud. Sci. Nat.*" And what are we supposed to do with a list of 45 references to phylogenetically based species concepts? It is as if a Java Web-bot has been set loose to gorge itself on the web of science with a few key words for guidance and has then emptied itself in the privacy of the book's back pages in one giant, ghastly, purgative dump.

Many years ago there was a *Punch* cartoon in which a bishop apologizes to his timid curate for giving him a bad egg at breakfast. Not wanting to give offence, the curate replies: "Oh, no, my Lord, I assure you that parts of it are excellent!" I find myself in the position of the curate: this book is much like the curate's egg, in the original sense of the phrase.

Sean Nee is in the Division of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK.



A splash of colour

The Great Barrier Reef is a living expanse of coral that stretches for over 2,000 kilometres along the east coast of Australia. It is home to an enormous variety of colourful fish, such as the

purple anthias shown above. A collection of stunning photographs of the reef by David Doubilet is now published in *Great Barrier Reef* (National Geographic Society, \$40).

A twist of fate

Helen M. Blau

It has long been known that sperm and eggs fuse together selectively, and not randomly with any other cell type. Some viruses, which are among the most primitive of organisms, express fusion proteins that allow them to penetrate mammalian cells. Furthermore, cells transformed by polyoma virus have been found to fuse spontaneously with other cells *in vivo*. Damaged muscle fibres — cells that each contain hundreds of nuclei — are repaired by the fusion into each cell of single-nucleus, undifferentiated cells (myoblasts). Thus, fusion has long been known to hold advantages for specialized cell types and primitive organisms. But it was a great surprise to learn that 'stem cells' derived from adults, rather than from developing tissue, might use fusion as a mechanism to change their fate.

Recent findings in tissue culture, showing that embryonic stem cells can fuse either with the precursors of neuronal cells or with adult bone-marrow cells, have initiated a new and important debate. As these experiments were carried out in culture with strong selective pressure, rather than under more natural *in vivo* conditions, some technical objections remain to be addressed. Nonetheless, these studies have spawned the new and important idea that fusion could be a significant means by which some cells might be 'reprogrammed' for a different function in adulthood.

One can envision several mechanisms that might cause adult bone-marrow-derived cells (stem cells) to change their morphology and activate new genes. The new work suggests that cell fusion could constitute one such mechanism. As a result, numerous earlier reports claiming that stem cells can activate previously silent genes, characteristic of a different cell type within diverse tissues, are now being questioned. Although it seems likely that some of the recently observed results reflect completely new changes in the fate of cells originally destined for some other function, others may merely be due to cell fusion.

But is there also something profound and exciting about the idea of 'mere' fusion? It seems quite possible that fusion of adult bone-marrow-derived cells may have an unexpected function, such as preventing cells from dying by providing a healthy and entire genetic complement. For instance, one might imagine that the highly specialized cells of tissues such as liver, brain and muscle would be difficult to form *de novo* in an adult, and might therefore be more amenable to 'rescue' through fusion.

Fusion of bone-marrow-derived cells might also serve as a means of supplying cells with new genes, such as tumour-suppressor genes, or of correcting genetically defective cells. If this proves to be possible, how exciting it will be to think that this might be a repair mechanism that continues to function throughout our lives and was also active in the lives of our evolutionary predecessors. It is possible that much more damage occurs in differentiated tissues than we imagined, because bone-marrow-derived cells gain access to the injured cells and restore them — at least up to a point. Beyond that point, when the balance between damage and repair shifts and fusion cannot cope with the demand, disease ensues.

Cell fusion has long been known to achieve effective reprogramming of cells. Almost two decades ago, my laboratory produced stable 'heterokaryons' by fusing specialized human cells of all three lineages with mouse skeletal-muscle cells to determine whether differentiation is irreversible. 'Terminal differentiation' implies a finality of fate that was, at that time, generally accepted. However, in the non-dividing multinucleate heterokaryons we produced, muscle genes were activated in primary human diploid keratinocytes, fibroblasts and hepatocytes that had no need of those genes.

Since then, others have replicated these findings in heterokaryons, using similar and different cell types. We went on to show that this reactivation of previously silent genes

Stem-cell fusion

The possibility has emerged that fusion could be a significant means by which cells might be 'reprogrammed' for a different function in adulthood.

occurs in the absence of DNA replication. Gene dosage, or the balance of proteins derived from the two cell types, determines which genes are activated. These results overturned the dogma that differentiation is fixed and irreversible, and showed that the phenotype of a differentiated cell is plastic, dynamic and determined by the balance of regulators at any given time.

These discoveries of genetic reprogramming in different cell systems redefined how differentiation was viewed — as an active rather than passive process. The exciting possibility that heterokaryons can exist *in vivo* is an extension of these findings that may soon be demonstrated in reality. Understanding the underlying mechanism of this process would not only be of great fundamental interest, but also should allow increased efficiency of fusion, which could lead, for example, to the rescue of muscle cells in muscular dystrophies, diseases for which there still is no cure, despite the fact that for some types of dystrophy a gene has been cloned.

Once the signals and underlying mechanisms are better understood, adult bone-marrow-derived cells could be genetically engineered to deliver genes to specific targets, giving rise to a new form of gene therapy. The prospect of heterokaryons that occur naturally in humans is therefore a matter of wonder, a feast for the imagination. The fusion that takes place when sperm meets egg, the beginning of organismic life, may also underlie a process of repair in adulthood that helps to maintain life. As the poet Robert Browning wrote: "Come grow old with me! The best is yet to be."

Helen M. Blau is director of the Baxter Laboratory for Genetic Pharmacology, Stanford University, Stanford, California 94305, USA.

FURTHER READING

- Fenyo, E. M., Wiener, F., Klein, G. & Harris, H. *J. Natl Cancer Inst.* **51**, 1865–1875 (1973).
 Wiener, F., Fenyo, E. M., Klein, G. & Harris, H. *Nat. New Biol.* **238**, 155–159 (1972).
www.nature.com/nature/stemcells
 Blau, H. M. *Annu. Rev. Biochem.* **61**, 1213–1230 (1992).
 Blau, H. M. *et al. Cell* **105**, 829–841 (2001).

P. MOTTA/UNIV. LA SAPIENZA/SPL



Could cells from bone marrow (brown), such as blood cells, be engineered for use in gene therapy?

Cold antihydrogen

Tom W. Hijmans

The production of antihydrogen, the antimatter equivalent of the hydrogen atom, at low temperatures is an impressive feat. It also raises the possibility of searching for fundamental differences between matter and antimatter.

On page 456 of this issue, scientists working at the European Centre for Nuclear Research (CERN), in Geneva, report the first production of cold antihydrogen atoms¹. This is a crucial step towards an eagerly sought goal in physics — the realization of magnetically trapped neutral antimatter.

The ordinary matter that makes up ourselves and the world around us is based on elementary particles such as protons, neutrons and electrons. All of these particles have partners known as antiparticles. Antimatter is like a mirror image of ordinary matter, the most notable difference being that the electric charges of antiparticles are the opposite of those of their normal-matter counterparts. Matter and antimatter are deceptively similar: were we to live in an antimatter world, and be composed of antimatter ourselves, we would not be able to tell.

When matter and antimatter meet, however, dramatic things happen. When a particle collides with its antiparticle, the result is the complete annihilation of both. All that remains is a burst of radiation, the energy E of which is given by Einstein's famous equation $E = mc^2$, where m is the total mass of the particles involved in the collision and c is the speed of light. It is this complete transformation of mass into energy that has made antimatter a firm favourite in the world of science fiction.

Many physicists believe that the study of neutral antimatter, such as the antihydrogen atom, holds the key to questions of fundamental importance in physics — in particular, whether there is some fundamental difference between matter and antimatter, and the mystery of why the Universe appears to be overwhelmingly filled with matter. The obvious challenge for experimentalists is that their laboratories and measuring instruments are made of ordinary matter, and antiparticles will self-destruct on first contact with the apparatus.

The proposed solution is 'wall-free' magnetic confinement or trapping. The idea is to use an arrangement of magnetic fields to push

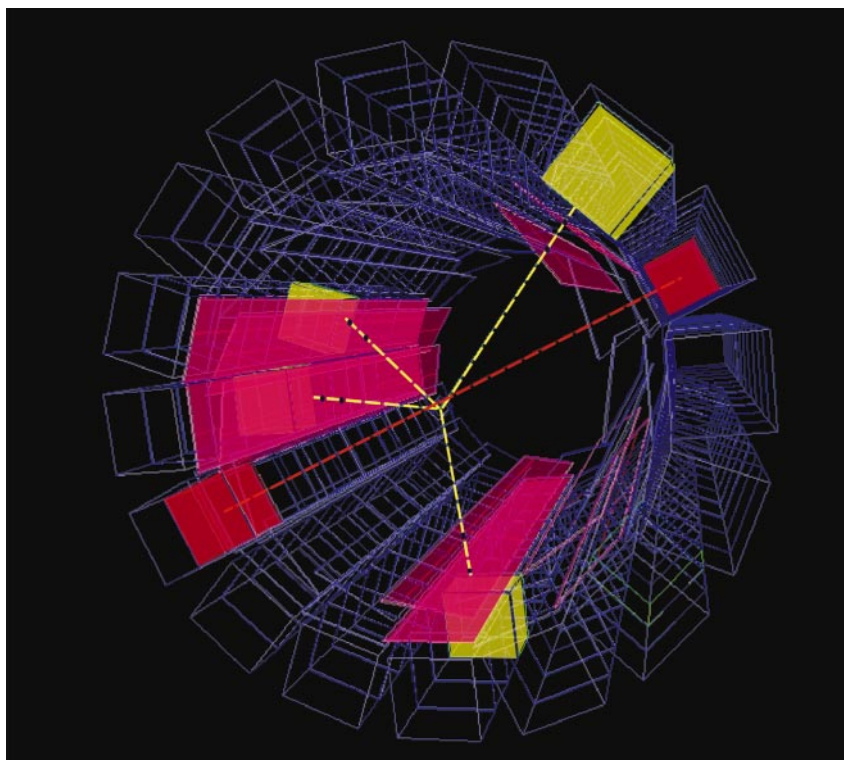


Figure 1 Antihydrogen annihilation. The ATHENA experiment has produced the first cold anti-atoms¹, detected through their destruction in collisions with matter particles. The annihilation of the antiproton produces secondary particles, which, picked up in the surrounding detectors (blue), can be traced back (yellow dashed lines) to the annihilation point. Similarly, the annihilation of the positron produces a distinctive back-to-back two-photon signature (red dashed lines). The overlap of the two annihilation points signifies that the positron and antiproton were bound together in an atom of antihydrogen.

the anti-atoms away from the walls of the apparatus and keep them suspended in space. Non-destructive precision experiments could then be carried out on the anti-atoms, using laser light, for example. Unfortunately, the magnetic forces that can be exerted on electrically neutral objects such as anti-atoms are rather limited. Hence the need for low-temperature experiments: unless the atoms are moving at the very low speeds characteristic of temperatures just a few degrees above absolute zero, they will simply pierce the magnetic confinement and hit the walls.

It is sometimes argued that the considerable effort needed to make antihydrogen can only be justified by the proposed experiments that will seek out possible differences between matter and antimatter. There are sceptics who express reservations about such

experiments, pointing to the high likelihood that no differences will be found. All agree, however, that if any difference were found, no matter how small, it would cause a big stir in physics. Personally, I consider the stable confinement of neutral antimatter to be valuable in itself. Just to be able to look inside this untouchable realm that is similar to, but at the same time incompatible with, the world we live in has great aesthetic appeal.

Antihydrogen is the simplest possible antimatter atom: it consists of a positively charged positron, the antiparticle of the familiar electron, orbiting an antiproton that has negative electric charge. At CERN, two international collaborations, ATRAP and ATHENA, have for several years been inching towards the synthesis of antihydrogen from its charged constituents at low temperature. An

Further News and Views coverage appears on pages 493–497, with three articles discussing papers on the genome sequences of malaria parasites, and of the mosquito that is the principal vector of the disease in humans.

essential innovation introduced by ATRAP was the nested Penning trap², a clever device that solves the problem of how to confine oppositely charged positrons and antiprotons simultaneously. ATRAP came quite close to producing antihydrogen last year³ but it is ATHENA that now reports success¹.

The antiprotons supplied to ATRAP and ATHENA from a particle accelerator move at almost the speed of light. First, this velocity must be reduced by an enormous factor, and during this process many particles are lost. The other constituents, the positrons, are emitted by a radioactive source and must also be slowed down to be useful. Having as many particles as possible to begin with is the key to success. The formation of an anti-atom in collisions between antiparticles is a rare event and, even when formed, most anti-atoms are not detected: about 130 antihydrogen atoms were observed in the ATHENA experiment, out of an estimated 50,000 produced.

Unfortunately, it is not yet possible to trap the anti-atoms. In fact, they reveal their presence only through their destruction in collisions with the apparatus walls (Fig. 1). An annihilating antiproton produces a number of energetic particles that fly off in various directions and these secondary particles leave directional traces in the detectors that surround the anti-atom sample. Like a ballistics expert on the scene of a crime who recon-

structs the position of a gunman by tracing back trajectories from scattered bullet holes, the physicists are able to determine precisely the location of the antiproton annihilation.

It is then necessary to prove that the antiproton was not an isolated particle, but part of an antihydrogen atom. To show this, it is sufficient to find a positron that has also been annihilated at the same position and precisely the same time. The tell-tale signature of such a positron–electron annihilation in the wall material is the emission of two γ -ray particles of well-defined energy (512 keV each) that leave the ‘scene of the crime’ in exactly opposite directions.

The production of cold antihydrogen is a milestone, but it is just the beginning. Taking the next step towards trapping anti-atoms is fraught with difficulties. For one thing, the arrangement of magnetic fields needed to create a nested Penning trap is not optimal for trapping neutral anti-atoms. But the field is buzzing with ideas and ATHENA’s progress will surely provide the motivation for researchers to face these challenges. ■

Tom W. Hijmans is at the Van der Waals–Zeeman Instituut, Valckenierstraat 65/67, 1018 XE Amsterdam, The Netherlands.
e-mail: hijmans@science.uva.nl

1. Amoretti, M. *et al.* *Nature* **419**, 456–459 (2002); advance online publication 18 September 2002 (doi:10.1038/nature01096).
2. Gabrielse, G. *et al.* *Phys. Lett. A* **129**, 38–42 (1988).
3. Gabrielse, G. *et al.* *Phys. Lett. B* **507**, 1–6 (2001).

Evolutionary biology

Death in the slow lane

Marcel Cardillo and Adrian Lister

Were the Late Pleistocene extinctions of large mammals the result of climate change or big-game hunting by humans? Reconstructing the biology of extinct species provides clues to the answer.

What caused the Late Pleistocene ‘megafaunal’ extinctions — the episode between about 50,000 and 10,000 years ago when mammoths, giant ground sloths, giant kangaroos (Fig. 1) and dozens of other large vertebrate species became extinct? The ‘overkill’ theory holds that human hunters drove the megafauna to extinction. An extreme form of the overkill theory is the ‘blitzkrieg’ model, in which the humans of Pleistocene times were big-game hunters, selectively and rapidly hunting the largest species to extinction as they swept through newly colonized continents. Others argue that the megafauna were killed off by climate and vegetation change at the end of the last ice age.

Usually, these theories are assessed by using the fossil record to compare the timing of megafaunal extinctions with human arrival on continents and climate change. But as he describes in a paper published in *Proceedings of the Royal Society*¹, Chris

Johnson has taken a fresh approach by systematically comparing the traits of extinct mammal species with those that survived. He finds that it was not large size that predisposed species to extinction, but low reproductive rate. This does not fit in with the blitzkrieg view of events.

Johnson models the close relationship between body size and reproductive rate within taxonomic families of living mammals, and uses this to infer reproductive rates for extinct members of the same families. He cleverly uses a between-family comparison to sidestep the potential circularity of this method, and then shows that the likelihood of extinction was higher for groups with lower reproductive rates, regardless of their body size. Even relatively small mammals became extinct if their fecundity was low enough. Strikingly, the threshold reproductive rate at which the chance of extinction exceeds 50% is roughly the same — about one offspring per female per year — for all

groups examined, from 700-gram lemurs to one-tonne cattle. It seems that what was lost during the Pleistocene was not so much the megafauna as the ‘bradyfauna’ — a whole way of life based on slow life-history.

These results run counter to the idea of extinction by rapid blitzkrieg because, if larger mammals were being selectively hunted, body size as well as reproductive rate should be an important determinant of extinction. However, Johnson noticed another interesting pattern in his data. Those mammal species that bucked the trend, surviving to the present day despite low reproductive rates, tend to be arboreal, nocturnal or inhabitants of dense forests, high latitudes or high altitudes — all of which ought to protect them from human hunters. Johnson interprets this as evidence in favour of more general overkill, where species of all body sizes were harvested.

Johnson’s findings complement those of Alroy², who modelled the effects of human hunting on Late Pleistocene extinctions in North America. Without assuming human preference for big game, Alroy’s model matched the observed pattern of large-mammal extinction. The two studies agree that even low levels of hunting could have led to extinction: Johnson suggests that, because of the slow reproductive rates of the victims, extinctions need not have occurred rapidly, while Alroy gives a median extinction date in North America of about 900 years after the Clovis hunter–gatherers migrated into the continent from Asia about 13,400 years ago. Whether or not this can be considered rapid blitzkrieg is a matter of perspective — ecologists and palaeontologists are used to working at very different timescales.

Johnson’s result leaves the effects of climate change an open question. Species with slow reproductive rates would have been the most vulnerable to environmental degradation as well as to hunting. For Australia, it has been suggested that during the Last Glacial Maximum (around 20,000 years ago), expansion of the arid interior at the expense of the lush coastal zone was a causal factor in Pleistocene extinctions. The finding that, even in Australia, survivors tended to live in forest or other cryptic habitats, could be interpreted as evidence against this climate model. Moreover, some dates for Australian fossil sites^{3,4} place megafaunal extinctions well before the Last Glacial Maximum, at around the time of the first evidence of humans.

Perhaps a similar approach to Johnson’s could be used to test predictions of the climate hypothesis — comparing, say, extinction patterns in Australian taxa such as reptiles and mammals that differ in their ability to cope with extreme drought. Further, the climatic models for other continents are very different and need to be tested separately. The disappearance of the



Figure 1 Prehistoric victim. A reconstruction of *Procoptodon*, a giant short-faced kangaroo from the Late Pleistocene of Australia, which stood about 3 m high.

steppe–tundra and other mixed vegetational environments in the northern continents, to be replaced by zoned forest, has been implicated in extinctions of mammoth, woolly rhino and other species there⁵. Here, the patterns of survival may fit climate/vegetational models at least as well as those invoking overkill.

The Pleistocene extinctions did not touch all regions of the world. In southern Asia, and especially Africa, the megafauna survived largely intact through the Pleistocene to the present. Africa was excluded from Johnson's study, but why did it suffer so few extinctions? Overkill theorists claim that, because humans originated there, African mammals coevolved with people and were thus less 'naïve' to human predation. Climate theorists, on the other hand, propose that the array of African vegetation types was modified relatively little by Pleistocene climate change, allowing the survival of even the largest mammals (elephants, rhinos, and so on). Johnson's approach, centred on taxonomic groups, points to a different line of enquiry. Could it be that African mammals, several groups of which are believed to be related and which are collectively known as the Afrotheria⁶, tend to have faster life-

history relative to body size than mammals of other continents?

We can draw parallels with the current extinction crisis from Johnson's findings. Alroy points out that the modelled rates of Late Pleistocene extinction were too slow to be perceived by the humans of the time. Today's extinctions have accelerated to an observable pace. Moreover, slow life-history is a strong predictor of current extinction risk in living mammals⁷. Perhaps in another 50,000 years — or even sooner — we will be left only with those that live life in the fast lane. ■

Marcel Cardillo is in the Department of Biological Sciences, Imperial College, Silwood Park, Ascot SL5 7PY, UK.

e-mail: m.cardillo@ic.ac.uk

Adrian Lister is in the Department of Biology, University College London, London WC1E 6BT, UK. e-mail: a.lister@ucl.ac.uk

1. Johnson, C. N. *Proc. R. Soc. Lond. B* published online 23 September 2002 (doi: 10.1098/rspb.2002.2130).
2. Alroy, J. *Science* **292**, 1893–1896 (2001).
3. Roberts, R. G. *et al. Science* **292**, 1888–1892 (2001).
4. Miller, G. H. *et al. Science* **283**, 205–208 (1999).
5. Guthrie, R. D. in *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin, P. S. & Klein, R. G.) 259–298 (Univ. Arizona Press, Tucson, 1984).
6. Murphy, W. J. *et al. Science* **294**, 2348–2351 (2001).
7. Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. *Proc. R. Soc. Lond. B* **267**, 1947–1952 (2000).

NATURE

100 YEARS AGO

So much has been written of late on totemism that I feel some diffidence in burdening still further the literature of the subject. But I may plead a slight claim on your attention, as I happen to be an unworthy member of the Crocodile kin of the Western tribe of Torres Straits, and I have been recognised as such in another island than the one where I changed names with Maino, the chief of Tutu, and thereby became a member of his kin. ... What is most needed at the present time is fresh investigation in the field. Those who are familiar with the literature of the subject are only too well aware of the imperfection of the available records. There are several reasons which account for this. Some of the customs and beliefs associated with totemism have a sacred significance, and the average savage is too reverent to speak lightly of what touches him so deeply. Natives cannot explain their mysteries any more than the adherents of more civilised religions can explain theirs. Further, they particularly dislike the unsympathetic attitude of most inquirers, and nothing shuts up a native more effectually than the fear of ridicule. From *Nature* 2 October 1902.

50 YEARS AGO

The attitude of the general science graduate to experiments involving subjective judgments is curious and illogical, even if understandable. He is taught throughout his study period to believe that those things which he measures during the course of his 'practical work' are facts, inviolable and true. ... Above all, he believes in the dogma of Kelvin, that we must measure to be able to understand. He rarely succeeds in grasping the principles of uncertainty in physical measurement, and many years may pass before he realizes that all measurement demands a judgment on the compromise between accuracy and expediency. As a result, he dismisses all experiments which involve direct subjective judgements as being 'vague and inaccurate'. This dogmatic attitude is not confined to the young graduate. It finds its way into higher levels, where great efforts are made to develop physical analogues to supplant the human observer in a perceptual situation. Consequently, it happens that the experimenter who uses human subjects as indicators finds it necessary to go to unusual lengths to explain his means as well as his ends. From *Nature* 4 October 1952.

Immunology

Survival of the fitters

Hans-Georg Rammensee

Our immune system's ability to fight viruses inside our cells relies on viral protein fragments being displayed on the cell surface. The enzyme needed for the last step in this presentation process has now been discovered.

Each cell in our body is covered with around 10,000 tiny protein fragments, representing almost every single protein being made in the cell. These fragments are peptides made up of exactly nine amino acids; they are glued to peptide receptors and convey information to the immune system about the interior of the cell. On page 480 of this issue, Serwold and colleagues¹ describe the enzyme that is responsible for producing fragments of this precise size. It chips away at larger peptides, removing one amino acid at a time, until they are either completely destroyed or are saved because they fit into a waiting peptide receptor inside the cell.

These peptide receptors are technically known as major histocompatibility complex (MHC) class I molecules, and they are essential to our body's ability to detect intracellular intruders. Killer T cells of the immune system screen cell surfaces for MHC-bound peptides that do not belong. As the 10,000 or so different surface peptides represent virtually all cellular proteins — each protein contributes one or a few snippets — the killer T cells can 'see' whether, for example, a viral protein is present in a screened cell. If so, they will eventually take action to kill that cell.

The peptides presented are sampled from the cellular waste. This process is called antigen processing and, with the exception of the last step (which is now described by Serwold *et al.*¹), is well documented (Fig. 1). Sooner or later, every cell protein is subject to degradation and the subsequent recycling of its amino acids. Old and worn-out proteins, as well as freshly made but faulty or incomplete proteins, are shredded by the proteasome — a barrel-shaped cellular machine. The proteasome draws proteins into its interior and cuts them into discrete pieces of up to 15 amino acids.

One end of these peptides, the carboxy-terminal end, is of a kind (hydrophobic or charged) that is looked on favourably by MHC class I molecules. The other end, the amino terminus, can vary greatly. The peptides are released into the surrounding cytosol and are attacked by aminopeptidase enzymes, which remove amino acids from the amino terminus. Most peptides are completely destroyed in the cytosol, but some manage to reach the membrane surrounding another cellular compartment, the endoplasmic reticulum (ER). They pass into the ER through an energy-consuming toll-gate, the transporter associated with antigen

processing (TAP) protein, which selectively allows peptides of 8 to 15 amino acids through. Inside the ER, newly synthesized, empty MHC molecules are waiting to be loaded with peptides.

Here we face a problem that first needs some background explanation. The antigen-processing machinery described so far is the same in every human being, and is even essentially the same in mice and humans. But each individual is different in regard to antigen processing inside the ER. The genes encoding MHC molecules vary greatly: there are several hundred different forms in the human population. So the products of these genes, the peptide receptors, are functionally very different. Although all MHC class I molecules bind peptides of nine amino acids, their individual binding specificities are dependent on peptide sequence and vary tremendously. As a result, different individuals present different sets of 10,000 peptides on their cells, even if the proteins inside the cells are identical. This leads to a marked individualization of the immune response, which explains people's differing sensitivities to infections and autoimmune diseases. This is

also the reason why our bodies tend to reject organs transplanted from another person.

So the problem is: how can the conserved processing machinery outside the ER turn out the correct peptides for the diverse MHC molecules lying in wait within? The answer is, it doesn't. The right peptides are instead produced by a random process driven by selection, as postulated many years ago². Specifically, as Serwold *et al.*¹ describe, the 8–15-amino-acid peptides entering the ER are attacked by the aminopeptidase ERAAP, and are only saved from complete destruction if they precisely fit one of the empty MHC class I molecules. Cellular aminopeptidases able to trim longer peptides down to nine amino acids have been identified before^{3,4}, and aminopeptidase activity in the ER has been shown to be essential to produce certain MHC-bound peptides^{5,6}. It is the achievement of Serwold *et al.*, however, to have identified the gene encoding this aminopeptidase; to have shown its physical location in the cell; and to have demonstrated its significance in loading up MHC class I molecules.

Serwold *et al.* first produced extracts of ERs from the livers and spleens of 50 mice. They isolated an aminopeptidase activity from this extract by chromatography, and, by mass spectrometry, determined the sequence of a protein band on the chromatograph that corresponded to this activity. The sequence showed similarities to previously described aminopeptidases from rats and humans. Serwold *et al.* chose to adopt a new name for this enzyme, ERAAP — for 'ER

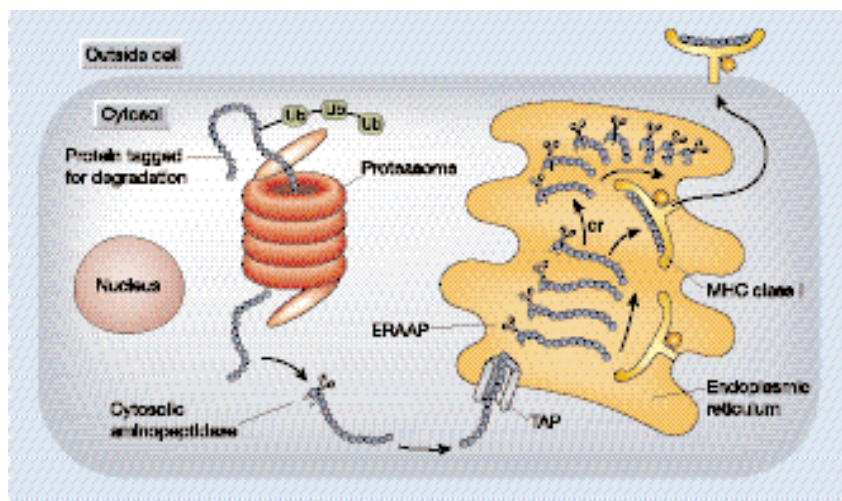


Figure 1 Informing the immune system. Sooner or later, every cellular protein reaches the end of its useful life and is degraded. This serves a beneficial purpose: nine-amino-acid peptides representing every cellular protein are taken to the cell surface, in complexes with MHC class I molecules, and presented to the immune system. The figure shows how this happens. First, a protein is marked for degradation with chains of the ubiquitin molecule (Ub), and fed into the cell's shredder — the proteasome — to be chopped into peptides of up to 15 amino acids. Aminopeptidase enzymes in the cytosol may further shorten these peptides, some of which then enter the endoplasmic reticulum via the TAP protein. There, as Serwold *et al.*¹ show, the peptides are attacked by the enzyme ERAAP and shortened one amino acid at a time until they are completely degraded — unless an intermediate nine-amino-acid peptide happens to fit into a waiting, empty MHC I molecule.

aminopeptidase associated with antigen processing' — to rhyme with TAP. The authors stained cells with an antiserum that binds to ERAAP, allowing them to detect it as an ER protein by microscopy.

As anticipated from earlier experiments³, Serwold *et al.* found that the expression of ERAAP was increased by treating cells with interferon- γ protein, which also upregulates other key elements in antigen processing. The authors also inhibited ERAAP expression, providing further insight into its importance for antigen processing: MHC class I expression was thereby reduced by about a third, and the presentation of different peptides was affected differently, as demonstrated in T-cell-recognition tests. Some MHC-presented peptides were almost unaffected by the absence of ERAAP, others disappeared, and two were actually presented to a greater degree than normal.

This wide range of effects nicely illustrates how MHC molecules become loaded with peptides — by predation and selection of the fittest, with the fittest being those peptides that are the best fitters. Some peptides happen to enter the ER as nine-amino-acid fragments; they can immediately bind to waiting MHC molecules and do not require ERAAP. Others are too long to bind, and require trimming by ERAAP before they can bind to MHC molecules. The percentage of long peptides seems to be considerable, as inhibiting ERAAP substantially decreases MHC expression on the cell surface (only MHC molecules loaded with peptides are stable). Still more peptides are the right size when they enter the ER, but are attacked by ERAAP and partially destroyed, leaving only a fraction that can reach the safety of the MHC molecules. All other peptides not able to bind MHC are completely destroyed by ERAAP, and cannot be found in extracts of whole cells².

What do we learn from this? First, the mechanism of antigen processing for MHC class I molecules — at least in this 'classical' pathway — is now fairly well understood from beginning to end. It is not immediately clear whether the molecular identification of ERAAP and its function will be of any practical benefit. One could envisage inhibitors of ERAAP being used to 'expose' a viral protein that is usually hidden from the immune system. Alternatively, such inhibitors might suppress the presentation of cellular peptides that usually activate the immune system in autoimmune diseases. As the entire immune system would be caught in a state of imbalance by inhibiting ERAAP, however, such ideas should be considered with great caution. It will certainly be interesting to see how the immune system works in ERAAP-knockout mice.

Second — and a more fundamental consideration — the way ERAAP works is a fine example of how nature uses the survival-of-the-fittest principle, even inside

the cell, to solve a complex task in an economical way.

Hans-Georg Rammensee is in the Department of Immunology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.
e-mail: rammensee@uni-tuebingen.de

1. Serwold, T. *et al.* *Nature* **419**, 480–483 (2002).

2. Falk, K., Rotzschke, O. & Rammensee, H.-G. *Nature* **348**, 248–251 (1990).

3. Beninga, J., Rock, K. L. & Goldberg, A. L. *J. Biol. Chem.* **273**, 18734–18742 (1998).

4. Stoltze, L. *et al.* *Nature Immunol.* **1**, 413–418 (2000).

5. Brouwenstijn, N., Serwold, T. & Shastri, N. *Immunity* **15**, 95–104 (2001).

6. Fruci, D., Niedermann, G., Butler, R. H. & van Endert, P. M. *Immunity* **15**, 467–476 (2001).

Earth science

Lava without the fizz

Peter Michael

The carbon dioxide content of lavas influences both the behaviour of Earth's mantle and its climate. In one particular case, the difficulties of directly measuring that content have been overcome.

Every day, hundreds of millions of tonnes of magma travel from Earth's interior to the surface, or close to it. Most of it crystallizes at shallow levels in the crust, the rest being ejected as lavas onto land or (more commonly) the sea floor. All of these magmas contain dissolved volatile molecules such as water, carbon dioxide, sulphur and chlorine, measurements of which could provide information about their concentrations and behaviour in the Earth's mantle. Such measurements would also allow calculation of the amount of volatiles entering the oceans, atmosphere and sediments (collectively known as the exosphere).

Carbon dioxide is of especial interest, because it can affect Earth's climate when released into the atmosphere. But it has hitherto proved impossible to measure directly: virtually every magma loses most of its dissolved volatiles as bubbles when it becomes 'oversaturated', owing to decreasing pressure as it ascends to the Earth's surface. On page 451 of this issue¹, however, Saal *et al.* report the first direct measurements of lava CO₂ that represent a magma's original composition. Their data permit the first reliable esti-

mations of the concentration and behaviour of CO₂ inside the mantle, and thus allow estimations of the amount of CO₂ supplied to the exosphere.

Magmas are the main pathway through which volatiles escape from the mantle to the exosphere, so the concentrations of their volatiles are of great interest to geologists. The basalt rocks that form at mid-ocean ridges (where the oceanic crust is being continuously produced by mantle melting) account for more than 85% of the Earth's magmatic production, suggesting that their contribution to the exosphere's volatile budget is significant. Magmatic volatiles become part of the atmosphere or oceans, and may eventually be incorporated into sedimentary rocks — as CO₂ in limestone or chloride in salt, for example. Volatiles return to the Earth's mantle when basaltic ocean crust that has been hydrated by sea water, and covered by sediments for millions of years, re-enters the mantle by the process of subduction.

To understand how the atmosphere and oceans formed and evolved, geochemists tabulate all of the inputs and outputs of volatiles to the exosphere. Such data can also help in

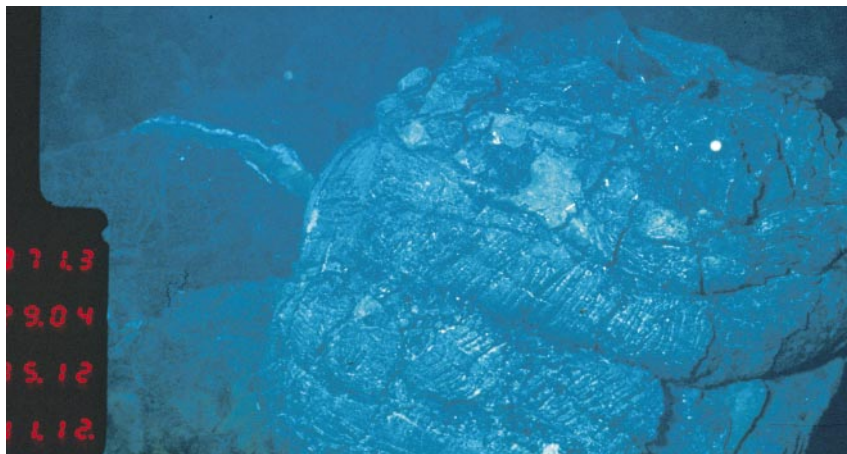


Figure 1 Sample site. A glassy lava formation on the Siqueiros fracture zone. Water depth here is 3,485 m, and the formation is about 3 m across. (Image courtesy D. J. Fornari, Woods Hole Oceanographic Inst., and M. R. Perfit, Univ. Florida. See also ref. 5.)

understanding the role of volatiles in the regulation of magma production², transport and explosiveness, as well as mantle viscosity³.

The process of volatile oversaturation in magmas is much like the effect of reducing pressure by opening a container of carbonated drink: as a magma rises, its volatiles come out of solution, forming buoyant bubbles that separate from the liquid and are to some degree lost. Geochemists have solved this problem for water and chlorine in mid-ocean-ridge basalts (MORBs) by examining submarine 'glasses' — rapidly cooled (quenched) liquids that were erupted under several hundred atmospheres of pressure on the sea floor, and retain all of these gases in solution. But the problem has remained for CO₂, because it typically becomes oversaturated even while the magma is still deep inside the Earth⁴. Seafloor pressures are usually insufficient to keep it in solution. (Similar problems still plague measurements of water, sulphur and chlorine in basalts that erupt at subduction zones.)

Saal *et al.*¹ got around the problem of oversaturation by analysing a peculiar glassy MORB (Fig. 1) from the Siqueiros fracture zone. This is a fault zone that separates long volcanic segments of the East Pacific Rise, part of the global mid-ocean-ridge system. They selected this particular lava because it has extremely low concentrations of elements such as potassium and niobium⁵, implying that it would also have very low initial concentrations of volatiles, as they tend to follow these elements into the liquid during mantle melting. It also had erupted at great water depth, so here was a chance to obtain a magma that had not crossed the threshold of volatile oversaturation. Besides measuring CO₂, the authors determined levels of sulphur, chlorine and water. These are of interest, but the real prize was the CO₂.

Saal *et al.* also analysed melt inclusions — tiny samples of liquid that became trapped at depth in crystals of olivine and were carried up by the magma. Most MORBs are thought to form by the coalescence of many small batches of liquid of diverse composition: some of them are rich in potassium, niobium and volatiles, others are depleted in these species. Melt inclusions are thought to capture this diversity before it is obliterated by mixing, but this idea is controversial. The CO₂ contents of the Siqueiros inclusions and host glass show nice correlations with elements such as potassium and niobium, trends that are usually destroyed by magma degassing.

Carbon dioxide content has been measured less directly in another special class of undersea glassy rocks^{6,7} — lavas that were so gas-rich that they were erupted and quenched, and so still contained myriad bubbles. The authors concerned assumed that there had been insufficient time for the lavas either to lose or to accumulate bubbles before they became frozen in glass, and they recon-

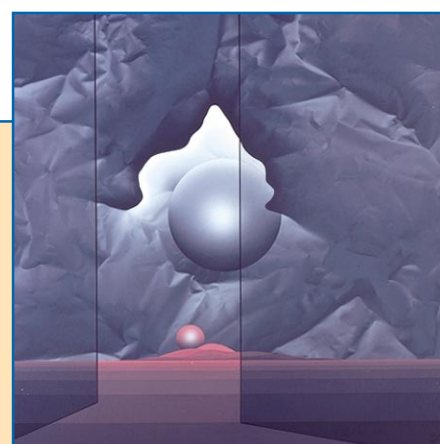
structed the original dissolved CO₂ content from measurements of CO₂ in glass and in bubbles. It is comforting to see that an extrapolation of Saal and colleagues' trends using a constant CO₂/Nb ratio projects to the CO₂ contents predicted in one of these papers⁶. So this ratio can potentially be used to predict the concentration of CO₂ in any MORB magma. From this point, it becomes possible to estimate the magma CO₂ output to the exosphere, both now and at particular times in the past. Scientists who model climate change can better evaluate whether intense volcanic activity in the past, such as the rapid creation of the mammoth Ontong Java Plateau 120 million years ago, could have supplied enough CO₂ to the atmosphere to drastically alter Earth's temperature⁸.

There is no guarantee that the CO₂/Nb relationship observed by Saal *et al.* will hold everywhere. In fact, deviations from that relationship would provide helpful clues about the heterogeneity in the mantle resulting from recycled subducted sediments, crust or primordial volatiles (as do variations in the water/cerium ratio in MORBs⁹). A more serious issue is whether Saal *et al.* are justified in using their measurements to calculate volatile concentrations in the upper-mantle source of MORBs, which is thought to be a major Earth reservoir. Is the Siqueiros lava, which is so highly depleted in melt-loving elements and volatiles compared to other MORBs, truly representative of this reservoir? The lava erupted in a fracture zone, not along a volcanic ridge segment, which is uncommon. It is possible to devise kinematic models in which lava forms in fracture zones by the remelting of recently melted mantle as it passes by its adjacent ridge segment. Determinations of volatile concentrations in additional rare, highly depleted glasses will help to confirm the ratios of CO₂/Nb and other volatiles, but the question will remain about how representative these samples are of the MORB reservoir.

Measurements in highly depleted melt inclusions from normal MORBs may help to resolve this issue. For the moment, however, Saal and colleagues' ratios provide the best way of estimating the original concentrations of CO₂ in magmas. ■

Peter Michael is in the Department of Geosciences, The University of Tulsa, 600 South College Avenue, Tulsa, Oklahoma 74104-3189, USA.
e-mail: pjm@utulsa.edu

1. Saal, A. E., Hauri, E. H., Langmuir, C. H. & Perfit, M. R. *Nature* 419, 451–455 (2002).
2. Gaetani, G. A. & Grove, T. L. *Contrib. Mineral. Petrol.* 131, 323–346 (1998).
3. Hirth, G. & Kohlstedt, D. *Earth Planet. Sci. Lett.* 144, 93–108 (1996).
4. Dixon, J. E., Stolper, E. & Delaney, J. R. *Earth Planet. Sci. Lett.* 90, 87–104 (1988).
5. Perfit, M. R. *et al.* *Earth Planet. Sci. Lett.* 141, 91–108 (1996).
6. Sarda, P. & Graham, D. *Earth Planet. Sci. Lett.* 97, 268–289 (1990).
7. Dixon, J. E. *et al.* *J. Petrol.* 38, 911–939 (1997).
8. Kerrick, D. M. *Rev. Geophys.* 39, 565–585 (2001).
9. Michael, P. J. *Earth Planet. Sci. Lett.* 131, 301–320 (1995).



Art

Imagine space

NASA's latest mission is somewhat unusual. The NASA Art Program has launched 'Copernica', a collection of specially commissioned artwork reflecting the wonder of the human adventure in space. But this online gallery is a little different: you can explore the collection interactively by navigating through a universe interpreted by artists. The images shown here are (from top): *Doorway to Space*, a "surrealistic view of the ever expanding vistas of space" by Patrice Breteau; *When Thoughts Turn Inward* by Henry Casselli, showing astronaut John Young waiting to board the space shuttle Columbia for its first launch in 1981; and Martin Berkon's *Jupiter Fly-by*, inspired by Voyager 2's images of the planet.

Alison Wright

Copernica is at

www.hq.nasa.gov/copernica/index.html



When the American sea sturgeon swam east

A colder Baltic Sea greeted this fish from across the Atlantic Ocean in the Middle Ages.

The two species of Atlantic sea sturgeon on either shore of the North Atlantic, *Acipenser sturio* in Europe and *A. oxyrinchus* in North America, probably diverged with the closure of the Tethys Sea and the onset of the North Atlantic Gyre 15–20 million years ago, and contact between them was then presumably precluded by geographic distance. Here we present genetic, morphological and archaeological evidence indicating that the North American sturgeon colonized the Baltic during the Middle Ages and replaced the native sturgeon there, before recently becoming extinct itself in Europe as a result of human activities. In addition to representing a unique transatlantic colonization event by a fish that swims upriver to spawn, our findings have important implications for projects aimed at restocking Baltic waters with the European sturgeon.

A. sturio once occurred abundantly in

rivers in regions from the Black Sea right up to the Baltic, but is now reduced to a relict population in southern France. *A. oxyrinchus*, however, is still found in rivers that run into the Atlantic from the Gulf of Mexico to Quebec.

We studied centuries-old museum specimens (*n*453; see supplementary information) and contemporary specimens of *A. sturio* (*n*467) and *A. oxyrinchus* (*n*4351) from almost all known populations. To investigate their molecular genetics, we compared the same two DNA fragments, one of about 200 base pairs from the mitochondrial D-loop¹ and the other of about 230 base pairs from nuclear DNA flanking the microsatellite Aox-23 (ref. 2).

Two *A. sturio* and 39 *A. oxyrinchus* mitochondrial DNA haplotypes were found that had 22 fixed differences between species. Fourteen archived specimens from the European Atlantic and the North, Adriatic

and Mediterranean seas showed the expected *A. sturio* haplotypes. But, surprisingly, ten archived specimens from the Baltic and one from the Oste River (North Sea) carried the *A. oxyrinchus* haplotype A (Fig. 1).

In the nuclear DNA, three consistent differences distinguished each species. Archived specimens with haplotype A carried *A. oxyrinchus* genotypes, whereas all other archived specimens had *A. sturio* genotypes. Nuclear sequences revealed no evidence of hybridization. Estimations based on a molecular clock³ indicate that the Baltic *A. oxyrinchus* originated from a single founding matrilineage (haplotype A) during the past 1,910 years (50% confidence interval).

Examination of the morphological differences associated with the two species^{4,5} revealed that archived Baltic sturgeon had a closer affinity with *A. oxyrinchus* than with *A. sturio*. In addition, archived Baltic specimens showed alveolar sculpting on their scutes — as seen in *A. oxyrinchus* — and not the tubercular-radial sculpting displayed by *A. sturio*⁵.

Re-examination of 972 scutes from 9 archaeological sites (Fig. 1b) indicated that *A. sturio* colonized Baltic waters some time after the Pleistocene (about 3,000 years ago), followed by *A. oxyrinchus* about 1,800 years later (see supplementary information). Both species were found in deposits in Ralswiek, Island of Rugia, suggesting that there had been a remarkable species shift from *A. sturio* to *A. oxyrinchus* between 1,200 and 800 years ago in the Baltic.

The subsequent decline of *A. sturio* and the establishment of a sustained *A. oxyrinchus* population in European waters during the Little Ice Age might have been due to selection for the hypothermal conditions that characterize their likely Canadian sources: *A. oxyrinchus* spawns between 13.3 and 17.8 °C (ref. 6), whereas *A. sturio* does not spawn below 20 °C (ref. 7).

Our findings have implications for restocking the Baltic Sea with sturgeon (HELCOM-Decision 18/97 project), an endeavour that has so far stagnated owing to the very limited availability of *A. sturio*. Attempts to reintroduce the North American *A. oxyrinchus* could also be hindered because this cold-water-tolerant species might no longer flourish in today's warmer Baltic waters, which would better suit the thermal tolerance of the European sturgeon.

Arne Ludwig*, Lutz Debus†, Dietmar Lieckfeldt*, Isaac Wirgin‡, Norbert Benecke§, Ingo Jenneckens||, Patrick Williot¶, John R. Waldman#, Christian Pitra*

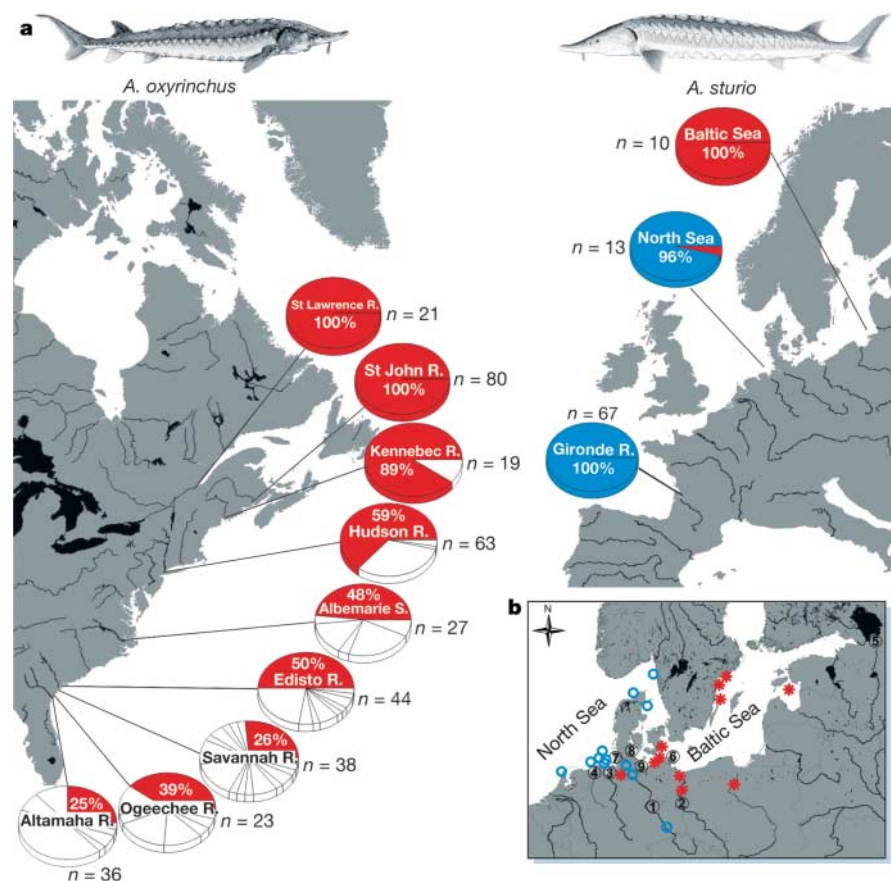


Figure 1 Geographical distribution of two lineages of mitochondrial DNA haplotypes found in Atlantic sea sturgeon from North America and Europe. **a**, Map (not to scale) showing sampling localities, number of sturgeon genetically analysed, and the distribution of *Acipenser oxyrinchus* haplotype A (red), all other *A. oxyrinchus* haplotypes (white) and *A. sturio* haplotypes (blue). The frequency of modal haplotype A increased with latitude among *A. oxyrinchus*. **b**, Localities of archival sampling sites (haplotypes: red, *A. oxyrinchus*; blue, *A. sturio*) and of archaeological data: 1, Elbe River (Niedergörsen); 2, Oder River (Lossow); 3, Weser River (Feddersen-Wierde); 4, Ems River (Jegum); 5, Lake Ladoga; 6, Ralswiek (Island of Rugia); 7, Eider River (Elisenhof); 8, Spey River (Schleswig); 9, Lübeck (Germany).

*Institute for Zoo and Wildlife Research,
Department of Evolutionary Genetics,
10252 Berlin, Germany
e-mail: ludwig@izw-berlin.de

†Institute of Zoology, University of Rostock,
18051 Rostock, Germany

‡Department of Environmental Medicine,
New York University School of Medicine, Tuxedo,
New York 10987, USA

§German Archaeological Institute,
Eurasian Division, 14195 Berlin, Germany

||Institute for Animal Breeding and Genetics,
University of Göttingen, 37075 Göttingen, Germany

¶Cemagref, Inland Living Aquatic Resources

Research Unit, 33612 Cestas cedex, France
#Hudson River Foundation for Science and
Environmental Research, New York,
New York 10011, USA

1. Ludwig, A., May, B., Debus, L. & Jenneken, I. *Genetics* **156**, 1933–1947 (2000).
2. King, T. L., Lubinski, B. A. & Spidle, A. P. *Conserv. Genet.* **2**, 103–119 (2001).
3. Brown, J. T., Beckenbach, A. T. & Smith, M. J. *Mol. Biol. Evol.* **10**, 326–341 (1993).
4. Magnin, E. *Le Naturaliste Canadien* **91**, 5–20 (1964).
5. Artyukhin, E. & Vecsei, P. J. *Appl. Ichthyol.* **15**, 35–37 (1999).
6. Borodin, N. *Trans. Am. Fish. Soc.* **55**, 184–190 (1925).
7. Quantz, H. *Mitt. Dt. Seefischerei-Vereins* **19**, 176–204 (1903).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Food chemistry

Acrylamide is formed in the Maillard reaction

Reports of the presence of acrylamide in a range of fried and oven-cooked foods^{1,2} have caused worldwide concern because this compound has been classified as probably carcinogenic in humans³. Here we show how acrylamide can be generated from food components during heat treatment as a result of the Maillard reaction between amino acids and reducing sugars. We find that asparagine, a major amino acid in potatoes and cereals, is a crucial participant in the production of acrylamide by this pathway.

Products of the Maillard reaction are responsible for much of the flavour and colour generated during baking and roasting. An important associated reaction is the Strecker degradation of amino acids by

these intermediates (Fig. 1), in which the amino acid is decarboxylated and deaminated to form an aldehyde.

We investigated whether this reaction could provide a possible route to acrylamide. The amino acid asparagine should be a particularly suitable reactant as it already has an amide group attached to a chain of two carbon atoms. We therefore performed a series of Maillard reactions between glucose and asparagine, as well as with other amino acids that do not have the correct carbon backbone for acrylamide (Fig. 1).

Significant quantities of acrylamide (221 mg per mol of amino acid) were found when an equimolar mixture of asparagine and glucose was reacted at 185 °C in phosphate buffer in a sealed glass tube. The temperature dependence of acrylamide formation from asparagine indicates that this is favoured above 100 °C and that very high temperatures are not necessary (Fig. 2). In similar reactions with glucose and glycine, cysteine or methionine at

185 °C, no acrylamide was detected (detection limit, 0.5 mg mol⁻¹). Glutamine and aspartic acid gave only trace quantities of acrylamide (0.5–1 mg mol⁻¹).

When a dry mixture of asparagine and glucose was reacted at 185 °C (that is, without buffer solution), only 25 mg mol⁻¹ acrylamide was formed. Although the dry reaction is a realistic system with which to simulate the later stages of baking and toasting of food, it is less efficient because the reactants are incompletely mixed in the absence of a solvent. Trace quantities of acrylamide were produced under these conditions from glutamine and aspartic acid, but not from any of the other amino acids apart from methionine, which yielded 5 mg mol⁻¹.

To test for the involvement of Strecker degradation in the the production of acrylamide, we used 2,3-butanedione instead of glucose in these reactions (butanedione is one of several dicarbonyl compounds formed in the Maillard reaction). Acrylamide was produced when asparagine was allowed to react with butanedione both in a dry system (40 mg mol⁻¹) and in buffer (63 mg mol⁻¹). Heating asparagine on its own at 185 °C did not produce acrylamide, confirming the requirement for the dicarbonyl reactant and Strecker degradation.

Again, there was no significant production of acrylamide in either system from butanedione and the other amino acids, with the exception of methionine (6 mg mol⁻¹ in the dry system). The Strecker aldehyde formed from methionine is methional, but acrolein can also be formed, together with ammonia: subsequent oxidation of acrolein to acrylic acid followed by amidation could then generate acrylamide (Fig. 1). However, this reaction might be limited by its requirement for ammonia, which reacts readily with carbonyls and other Maillard intermediates.

The almost exclusive formation of acrylamide from asparagine could explain the occurrence of acrylamide in cooked plant-based foods, such as cereals and potato, which are rich in this particular amino acid⁴. In potato used for the manufacture of potato crisps, the dominant free amino acid is asparagine (940 mg kg⁻¹, representing 40% of the total amino-acid content⁵); in wheat flour it is present at 167 mg kg⁻¹, corresponding to 14% of the total free amino acids (our unpublished results), and a high-protein rye variety contains 173 mg kg⁻¹ (18% of the total free amino acids)⁶.

Our findings indicate that Maillard reactions involving asparagine can produce acrylamide and might explain the increased concentrations of acrylamide in certain plant-derived foods after cooking.

Donald S. Mottram*, Bronislaw L. Wedzicha†, Andrew T. Dodson*

*School of Food Biosciences, The University of Reading, Whiteknights, Reading RG6 6AP, UK

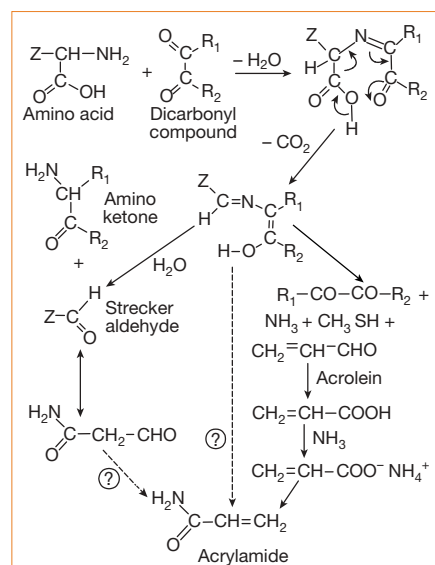


Figure 1 Proposed pathways for the formation of acrylamide after Strecker degradation of the amino acids asparagine and methionine in the presence of dicarbonyl products from the Maillard reaction. In asparagine, the side chain Z is $-\text{CH}_2\text{CONH}_2$; in methionine, it is $-\text{CH}_2\text{CH}_2\text{SCH}_3$.

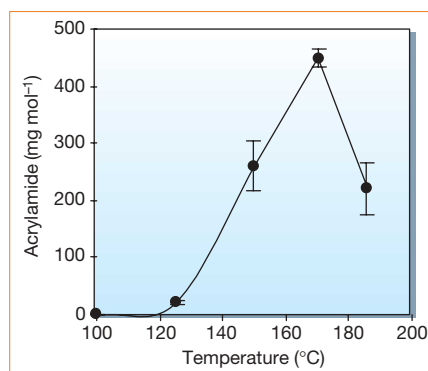


Figure 2 Temperature-dependent formation of acrylamide (mg per mol of amino acid) from asparagine (0.1 mmol) and glucose (0.1 mmol) in 0.5 M phosphate buffer (100 ml, pH 5.5) heated in a sealed glass tube for 20 min. Error bars represent standard deviations ($n=3$). Acrylamide produced in the reaction was extracted with ethyl acetate and analysed by gas chromatography with mass spectrometry after derivatization to 2,3-dibromopropanamide⁷, using 2-methylacrylamide as the internal standard. Selected ion monitoring was used to detect the analytes, with m/z 150 and 152 for acrylamide and m/z 120 and 122 for methylacrylamide. The presence of acrylamide in selected samples was confirmed in full mass spectra.

e-mail: d.s.mottram@reading.ac.uk

†Procter Department of Food Science, University of Leeds, Leeds LS2 9JT, UK

1. Rosen, J. & Hellenas, K.-E. *Analyst* **127**, 880–882 (2002).
2. Tareke, E., Rydberg, P., Karlsson, P., Eriksson, S. & Törnqvist, M. *J. Agric. Food Chem.* **50**, 4998–5006 (2002).
3. IARC IARC Monographs on the Evaluation of Carcinogenic Risks to Humans **60**, 389 (1994).
4. Belitz, H.-D. & Grosch, W. *Food Chemistry* (Springer, New York, 1999).
5. Martin, F. L. & Ames, J. M. *J. Agric. Food Chem.* **49**, 3885–3892 (2001).
6. Dembinski, E. & Bany, S. *J. Plant Physiol.* **138**, 494–496 (1991).
7. Castle, L. *J. Agric. Food Chem.* **41**, 1261–1263 (1993).

Competing financial interests: declared none.

Food chemistry

Acrylamide from Maillard reaction products

The discovery of the adventitious formation of the potential cancer-causing agent acrylamide in a variety of foods during cooking^{1,2} has raised much concern^{3,4}, but the chemical mechanism(s) governing its production are unclear. Here we show that acrylamide can be released by the thermal treatment of certain amino acids (asparagine, for example), particularly in combination with reducing sugars, and of early Maillard reaction products (*N*-glycosides)⁵. Our findings indicate that the Maillard-driven generation of flavour and colour in thermally processed foods can — under particular conditions — be linked to the formation of acrylamide.

We heated 20 amino acids individually at 180 °C for 30 min and found that acrylamide is formed under these conditions from methionine and from asparagine (3.651.4 and 0.5650.05 µmol acrylamide per mol amino acid, respectively; all results are averages of *n*46 independent determinations unless stated otherwise).

When pyrolysed at 180 °C with an equimolar amount of glucose, asparagine in particular generates significant amounts of acrylamide, reaching an average of 368 m mol mol^{−1} after an incubation time (*t*_i) of 30 min. If asparagine monohydrate was used in the incubation or water was added to the reaction (0.05 ml) before thermolysis, the release of acrylamide was enhanced nearly threefold (9605210 m mol mol^{−1}), or over 1,700 times the amount formed from asparagine alone under the same conditions.

Reaction of methionine and glutamine with equimolar amounts of glucose at 180 °C also increased the formation of acrylamide, which occurred rapidly in each case (*t*_i45 min; Fig. 1a). Cysteine was found to liberate acrylamide after condensation with glucose (2.050.8 m mol mol^{−1} at *t*_i430 min and 180 °C).

Investigating the role of different carbohydrates in the formation of acrylamide, we

found that pyrolysing any of these amino acids (Asn, Gln, Met, Cys) with an equimolar amount of D-fructose, D-galactose, lactose or sucrose all led to a significant release of acrylamide, with comparable yields from each sugar. No acrylamide was detected when any of these carbohydrates was heated alone.

To test whether early Maillard products such as *N*-glycosides could be acrylamide precursors in thermal decomposition reactions, we measured the yields of acrylamide after pyrolysis (*t*_i420 min, 180 °C) of 0.2 mmol of four different *N*-glycosides (Fig. 1b). Yields were significant (in m mol per mol *N*-glycoside: compound 1, 1,3055323; 2, 1,4195278; 3, 1452.7; and 4, 8.151.5) and comparable to those released from the amino-acid and reducing-sugar precursors under the same conditions. Furthermore, compound 1 was confirmed as an intermediate in the asparagine/glucose reaction by high-resolution mass-spectrometric analysis of a methanol extract of the pyrolysate.

On the basis of structural considerations, asparagine or the *N*-glycosides 1 and 2 could be direct precursors of acrylamide under pyrolytic conditions. Condensation of asparagine with ¹³C₆-labelled glucose confirmed that the amino acid is the carbon source of acrylamide. Upon pyrolysis, formation of the corresponding *N*-glycoside probably facilitates the decarboxylation step and heterolytic cleavage of the nitrogen-carbon bond to liberate acrylamide (CH₂5CHCONH₂). Although decarboxylation is favoured at higher temperatures, the *N*-glycosidic bond seems to facilitate the deamination step.

Further evidence to support this pathway to acrylamide production is provided by the 98.6% incorporation of nitrogen-15 label into acrylamide after the pyrolysis of ¹⁵N-amide-labelled asparagine with glucose; there was no incorporation into acrylamide when ¹⁵N-α-amino-labelled asparagine was used in the same reaction. Results from similar isotope-labelling experiments (not shown) to determine the route of acrylamide formation from different *N*-glycosides produced by glucose pyrolysis with glutamine or methionine are less clear-cut, which suggests that other pathways (such as that for homolytic cleavage) might also lead to acrylamide.

The *N*-glycosidic bond is labile in the presence of water⁶ or under acidic and neutral pH conditions⁷, hydrolysing rapidly to the reducing sugar and amino acid. At higher pH, however, *N*-glycosides can be isolated as bimolecular complexes in the presence of polyvalent alkaline or transition-metal ions⁸. In food-processing systems that incorporate conditions of high temperature and water loss, *N*-glycoside formation could be favoured; when this condensation occurs between reducing sugars and certain amino acids, a direct pathway is opened up to

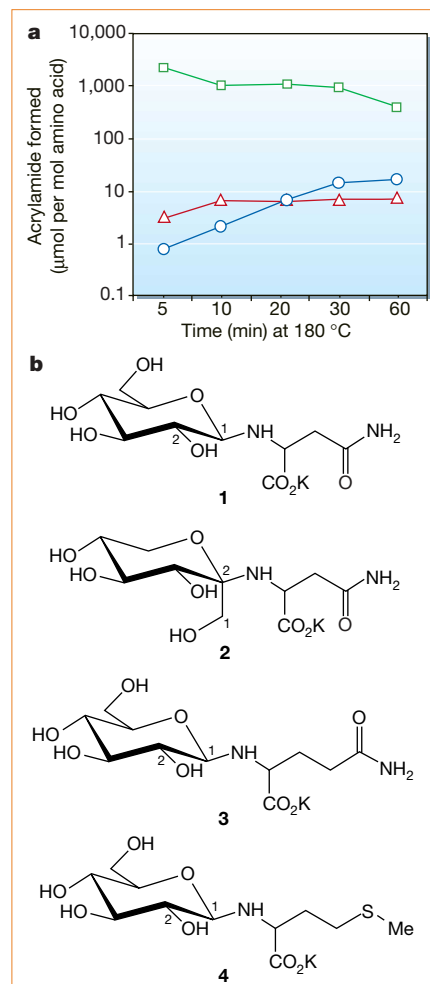


Figure 1 Production of acrylamide from *N*-glycosides. **a**, Logarithmic-scale plot of the formation of acrylamide over time in pyrolysates of glucose with glutamine (triangles), asparagine (squares) or methionine (circles). Each data point represents the average of *n*43 independent determinations; the coefficient of variation was less than 25%. For acrylamide analysis (by liquid chromatography coupled to electrospray ionization tandem mass spectrometry), pyrolysates were supplemented with ¹³C₃-acrylamide (50 ng), then suspended in hot water (more than 90 °C), sonicated and filtered before being applied to a solid-phase extraction cartridge (OASIS HLB, 0.2 g). Acrylamide eluted with 20% methanol was separated on a Shodex RSpak DE-613 polymer column with isocratic solvent flow. Detection by mass spectrometry was in the multiple-reaction monitoring mode with the characteristic fragmentation transitions for acrylamide (*m/z* 72→55, 72→27, 72→54) and confirmed by ion ratios (55/54 and 55/27). Further details are available from the authors. **b**, Chemical structures of the potassium salts of *N*-(p-glucos-1-yl)-L-asparagine (**1**), *N*-(p-fructos-2-yl)-L-asparagine (**2**), *N*-(p-glucos-1-yl)-L-glutamine (**3**) and *N*-(p-glucos-1-yl)-L-methionine (**4**).

potential progenitors of acrylamide.

Richard H. Stadler, Imre Blank, Natalia Varga, Fabien Robert, Jörg Hau, Philippe A. Guy, Marie-Claude Robert, Sonja Riediker
Nestlé Research Center, Nestec, Vers-chez-les-Blanc,
1000 Lausanne 26, Switzerland
e-mail: richard.stadler@rdls.nestle.com

1. Swedish National Food Agency website <http://www.slv.se>
2. Rosén, J. & Hellenäs, K.-E. *The Analyst* **127**, 880–882 (2002).
3. WHO FAO/WHO Consultation on the Health Implications of Acrylamide in Food (Geneva, 25–27 June 2002)

http://www.who.int/fsf/

4. European Commission Scientific Committee on Food (SCF) *Opinion of the Scientific Committee on Food on New Findings Regarding the Presence of Acrylamide in Food* (SCF/CS/CNTM/CONT/4 Final, 3 July 2002) http://europa.eu.int/comm/food/fs/scf/index_en.html
5. Ledl, F. & Schleicher, E. *Angew. Chem. Int. Ed. Engl.* **29**, 565–594 (1990).
6. Paulsen, H. & Pflughaupt, H. in *The Carbohydrates — Chemistry and Biochemistry* (eds Pigman, W. & Horton, D.) Vol. 1B, 881–927 (Academic, New York, 1980).
7. Von Euler, H. & Brunius, E. *Chem. Ber.* **59**, 1581–1585 (1926).
8. Chen, J., Pill, T. & Beck, W. *Z. Naturforsch. B* **44**, 459–464 (1989).

Competing financial interests: declared none.

Quantum cryptography

A step towards global key distribution

Large random bit-strings known as ‘keys’ are used to encode and decode sensitive data, and the secure distribution of these keys is essential to secure communications across the globe¹. Absolutely secure key exchange² between two sites has now been demonstrated over fibre³ and free-space^{4–6} optical links. Here we describe the secure exchange of keys over a free-space path of 23.4 kilometres between two mountains. This marks a step towards accomplishing key exchange with a near-Earth orbiting satellite and hence a global key-distribution system.

The security of our key-exchange system is guaranteed by encoding single photons using two sets of orthogonal polarizations. Our transmitter module (Alice; Fig. 1) incorporates a miniature source of polarization-coded faint pulses (approximating single photons; C.K., P.Z., M.H. and H.W., unpublished results), where 0° or 45° polarization encode binary zero, and 90° or 135° code binary one. These light pulses are expanded and collimated in a simple telescope to a beam of about 50 mm and then accurately aligned on the receiver (Bob; Fig. 1), a 25-cm-diameter commercial telescope. Light is collected and focused onto a compact four-detector photon-counting module (Fig. 1). A detection in any one detector then has an associated bit value, measurement basis (0° or 45°) and detection time. The bit values then form a raw key string. Valid bits are measured in the same basis as that in which they were encoded.

Alice and Bob use a standard communications channel, such as a mobile telephone, to ascertain which bits arrived (many are lost) and which measurement basis was used, then they both discard the invalid bits — which leaves them with nearly identical random bit-strings, the sifted key. Eavesdropping measurements on the single photons disturb the encoding and introduce errors of up to 25%, so Alice and Bob test for errors in a short section of sifted key to

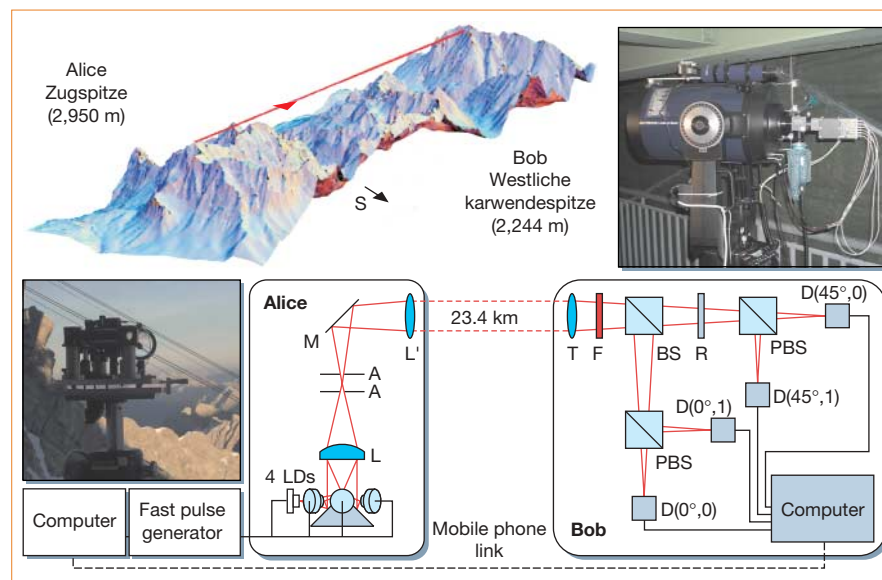


Figure 1 Overview of the experiment against a relief map of the trial site. In the Alice module, four separate lasers (LDs) encode the four polarizations based on a random bit-string fed from the Alice computer. They are combined in a spatial filter (A,A) using a conical mirror (M) and a lens (L). The beam expands to 50 mm and is collimated in an output lens (L8). In the Bob module, a telescope (T) collects the light, which is filtered (F) and then split in a polarization-insensitive beam-splitter (BS), passing on to polarizing beam-splitters (PBS) and four photon-counting detectors (D). One polarizing beam-splitter is preceded by a 45° polarization rotator (R). A click in one of the photon-counting detectors D(u, B) sets the bit value B and the measurement basis u.

verify the security of the channel. Low error rates due to background light detection and polarization settings are securely eliminated by using classical error-correcting codes sent over the mobile-telephone link.

In the long-range experiment, Alice was located at a small experimental facility on the summit of Zugspitze in southern Germany, and Bob was on the neighbouring mountain of Karwendelspitze, 23.4 km away. At this distance, the transmitted beam was 1–2 m in diameter and was only weakly broadened by air-turbulence effects at this altitude. Lumped optical losses of about 18–20 decibels were measured and, using faint pulses containing 0.1 photons per bit, the detected bit rate at Bob was 1.5–2 kilobits per second (receiver efficiency of 15%).

Operating at night with filters of 10-nm bandwidth reduced the background counts, and errors appeared in less than 5% of key bits. After sifting and error correction, net key exchange rates were hundreds of bits per second. In a series of experiments, several hundreds of kilobits of identical key string were generated at Alice and Bob.

In associated experiments in poorer visibility, we showed that key exchange could be carried out when transmission losses were up to 27 decibels, but improvements in receiver efficiency and background counts should take us beyond 33 decibels. With this performance, key exchange to near-Earth orbit (500–1,000 km range) should become possible.

Until now, the principal method of high-security key exchange has been the

‘trusted courier’ carrying a long random bit-string, the key, from one location to the other. Our experiment paves the way for the development of a secure global key-distribution network based on optical links to low-Earth-orbit satellites. We note that a 10-kilometre key-exchange experiment has recently been announced⁷.

C. Kurtsiefer*, P. Zarda*, M. Halder*, H. Weinfurter*, P. M. Gorman†, P. R. Tapster †, J. G. Rarity†

*Ludwig-Maximilians University, 80799 Munich, Germany

†Photonics Department, QinetiQ, Malvern, Worcestershire WR14 3PS, UK

e-mail: jgrarity@qinetiq.com

1. Singh, S. *The Code Book* (Anchor, New York, 1999).
2. Bennett, C. H. *et al.* *J. Cryptol.* **5**, 3–28 (1992).
3. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. *Rev. Mod. Phys.* **74**, 145–196 (2002).
4. Buttler, W. T. *et al.* *Phys. Rev. Lett.* **84**, 5652–5655 (2000).
5. Rarity, J. G., Gorman, P. M. & Tapster, P. R. *Electron. Lett.* **37**, 512–514 (2001).
6. Rarity, J. G., Gorman, P. M. & Tapster, P. R. *J. Mod. Opt.* **48**, 1887–1901 (2001).
7. Hughes, R. J., Nordholt, J. E., Derkacs, D. & Peterson, C. G. *New J. Phys.* **4**, 43.1–43.14 (2002).

Competing financial interests: declared none.

erratum

Cognitive change and the APOE 4 allele

I. J. Deary, M. C. Whiteman, A. Pattie, J. M. Starr, C. Hayward, A. F. Wright, A. Carothers, L. J. Whalley *Nature* **418**, 932 (2002)

In the second sentence of the seventh paragraph of this communication, the MMSE scores are incorrectly specified as less than or equal to 28; these should read as greater than or equal to 28.

Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle

Alberto E. Saal*, Erik H. Hauri†, Charles H. Langmuir* & Michael R. Perfit‡

* Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

† Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA

‡ Department of Geological Sciences, University of Florida, Gainesville, Florida 32611, USA

The analysis of volatiles in magmatic systems can be used to constrain the volatile content of the Earth's mantle and the influence that magmatic degassing has on the chemistry of the oceans and the atmosphere. But most volatile elements have very low solubilities in magmas at atmospheric pressure, and therefore virtually all erupted lavas are degassed and do not retain their primary volatile signatures. Here we report the undersaturated pre-eruptive volatile content for a suite of mid-ocean-ridge basalts from the Siqueiros intra-transform spreading centre. The undersaturation leads to correlations between volatiles and refractory trace elements that provide new constraints on volatile abundances and their behaviour in the upper mantle. Our data generate improved limits on the abundances of carbon dioxide, water, fluorine, sulphur and chlorine in the source of normal mid-ocean-ridge basalt. The incompatible behaviour of carbon dioxide, together with the CO_2/Nb and CO_2/Cl ratios, permit estimates of primitive carbon dioxide and chlorine to be made for degassed and chlorine-contaminated mid-ocean-ridge basalt magmas, and hence constrain degassing and contamination histories of mid-ocean ridges.

Studies of volatile elements in magmatic systems have particular importance, because volatiles influence mantle melting¹, magma crystallization and volcanic eruption, and their abundances and spatial distribution provide important constraints on models of mantle convection^{2,3}, mantle heterogeneity^{4–6} and crustal recycling^{7,8}. An advantage of studying mid-ocean-ridge basalts (MORBs) rather than sub-aerially erupted lavas is that the higher eruption pressures beneath the ocean inhibit degassing of volatile elements. Indeed, pioneering work on submarine lavas⁹ and the advent of improved analytical techniques have permitted more detailed and comprehensive studies of volatiles through investigations of mid-ocean-ridge magmatic systems^{10–21}.

Water, chlorine, sulphur and fluorine have sufficient solubility at sea-floor depths to allow an estimation of their initial concentrations in MORBs^{15–20}. But the solubility of CO_2 is so low that virtually all MORBs exsolve CO_2 -rich vapour as they approach the Earth's surface^{10–14}. As a result, nearly all MORB glasses are saturated with a CO_2 - H_2O vapour phase at the pressure of eruption, and frequently they are over-saturated with vapour. The super-saturation of CO_2 -rich vapour in MORB glasses results from a very rapid magma ascent rate that limits the time for bubble nucleation and growth, preventing complete CO_2 degassing at the pressure of eruption^{11,13}. There is no way of knowing how extensively the magmas have degassed, and therefore estimates of initial CO_2 concentrations, and abundances in the mantle from which they were derived, have varied by more than an order of magnitude^{7,8,22–26}.

Volatile content of Siqueiros melt inclusions

Melt inclusions (minute pieces of melt trapped in crystals) have been used in several recent studies to try to determine the pre-eruptive volatile contents of basaltic magmas^{4,27,28}. Melt inclusions provide an advantage over erupted magmas because they form before magma eruption, they are commonly trapped at pressures exceeding the pressure of eruption, and they are protected by their crystal host from eruptive degassing^{4,27}. In an effort to further constrain the pre-eruptive volatile content of MORB, we have analysed olivine-hosted primary melt inclusions in submarine

MORB from the Siqueiros transform fault, one of the most primitive sample suites on the East Pacific Rise (EPR)²⁹.

We investigated melt inclusions from seven Siqueiros samples: these samples were pillows from basaltic lavas erupted deep ($>3,500$ m) within the transform fault on the ocean floor. The inclusions were hosted by high-magnesium olivine phenocrysts (forsterite content 90–91%) sampled from the glassy rinds of the pillows. Most of the melt inclusions analysed are large (100–300 μm), glassy, with or without shrinkage bubbles, with no evidence of decrepitation or visible cracks. The glass inclusions are homogeneous, and have negative crystal shapes (the melt inclusion acquires the shape of the host crystal during entrapment) that indicate a primary origin³⁰. The major- and trace-element contents of the melt inclusions are exceptionally primitive, with slightly higher MgO content than the host glasses (Supplementary Information Tables 1 and 2, Supplementary Information Figs 5 and 6).

The Siqueiros melt inclusions and host glasses have low abundances of H_2O (0.037–0.122 wt%), CO_2 (44–244 p.p.m.), F (50–135 p.p.m.), S (495–1,024 p.p.m.) and Cl (1–21 p.p.m.; most melt inclusions have 1–3 p.p.m.) (see Supplementary Information Table 1). These compositions reach the lowest abundances so far measured for MORB, except for those that have been completely degassed because of eruption at hydrostatic pressure less than 50 bar (500 m water depth)^{9–20}.

The effect of eruptive degassing must be understood before the volatile content of the melt inclusions and host glasses can be used to investigate the variation of primary volatile content of basalts and their mantle sources. This is especially important for CO_2 , owing to its low solubility in basaltic magmas. The solubility model of Dixon *et al.*¹² permits evaluation of the H_2O - CO_2 saturation of the melt inclusions and host glasses at the depth of eruption. The Siqueiros glasses and melt inclusions range from nearly saturated to undersaturated in H_2O - CO_2 vapour at the pressures of eruption (Fig. 1). The very low (0–0.5 vol.%) vesicularity of the host glasses is consistent with the undersaturation of the magmas. The undersaturation at pressures of eruption is a unique feature of these data

that enables new constraints on the volatile content of their mantle source.

Comparison with non-volatile trace elements

Because none of the volatiles CO_2 , H_2O , Cl, F or S were degassed during eruption, their abundances can be compared to non-volatile trace elements whose mantle abundances and behaviour are well known. This allows us to establish the incompatibilities and mantle

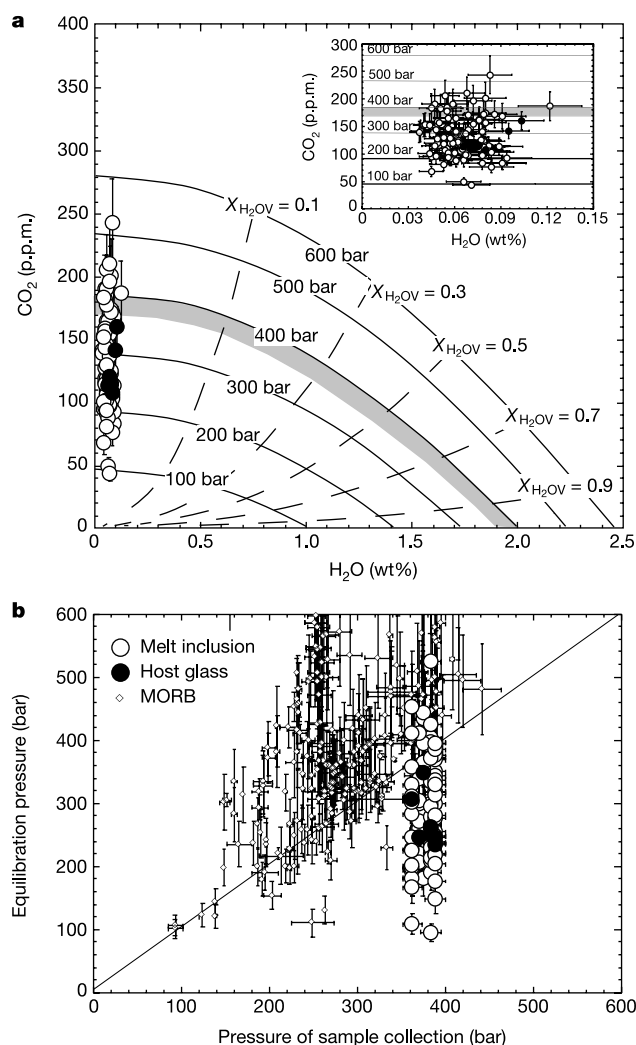


Figure 1 Pressure for CO_2 - H_2O saturation, and comparison with pressure of sample collection. **a**, CO_2 content versus H_2O content for Siqueiros glasses and melt inclusions. Solid lines, curves of constant pressure (isobars), according to the solubility model of ref. 12; dashed lines, curves of constant vapour composition. $X_{\text{H}_2\text{O}}^{\text{OV}}$ is the fraction of H_2O in the vapour phase in equilibrium with a basaltic melt at a given pressure. Grey area is the region between the 360-bar and the 400-bar isobars, corresponding to the depth of eruption of the Siqueiros basalts studied here. Inset, a magnified view of the data for Siqueiros glasses (filled circles) and melt inclusions (open circles). **b**, Equilibrium pressures calculated from dissolved CO_2 and H_2O concentration versus collection pressure. The calculated pressure was obtained using the model of ref. 12. Mid-ocean-ridge basalt (MORB) sample compilation (open diamonds) is from J. E. Dixon (personal communication) and A.E.S. (unpublished data). Error bars for Siqueiros glasses and melt inclusions are reported in Supplementary Information Table 1. H_2O - CO_2 saturated to supersaturated lavas plot on or above the 1:1 line (within error bars). The super-saturation of CO_2 -rich vapour in MORB glasses results from a very rapid magma ascent rate that limits the time for bubble nucleation and growth, preventing complete CO_2 degassing at the pressure of eruption. The few undersaturated MORB samples (apart from Siqueiros) were interpreted as having been transported down slope during eruption¹³. 1 bar = 10 m water depth.

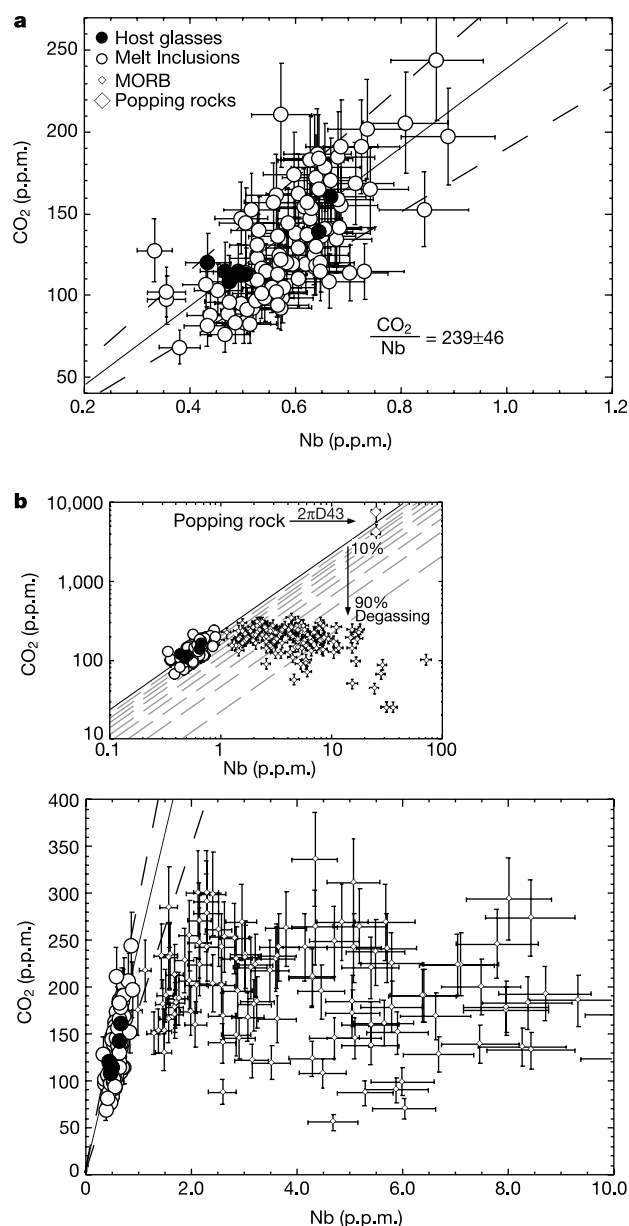


Figure 2 CO_2 contents versus Nb abundances in Siqueiros and other MORB samples. **a**, The CO_2 -undersaturation of many Siqueiros samples indicates that CO_2 was not degassed during eruption. Therefore, CO_2 concentrations will depend mainly on the behaviour of CO_2 during melting and the CO_2 content of the mantle source. The similar behaviour of CO_2 and Nb in the Siqueiros melt inclusions and glasses demonstrate the highly incompatible nature of CO_2 during melting. Thus, the CO_2/Nb ratios vary by only 20% (2σ , dashed lines) around an average value of 239 (solid line calculated as a simple regression). Using an estimate of the Nb concentrations in the MORB mantle (0.3 ± 0.05 ; refs 33, 34), we can constrain the CO_2 in the depleted mantle to be 72 ± 19 p.p.m. **b**, Bottom panel; Siqueiros glasses and melt inclusions have higher CO_2/Nb ratios than any other MORB samples. The CO_2 -undersaturation of the Siqueiros samples defines the CO_2/Nb ratios for undegassed MORB (solid line; 2σ , dashed lines). Top panel, the same data but using a logarithmic scale and including the values for popping rock (2TID43, the best example of undegassed MORB sample). We present two different CO_2 values for the popping rock: one measured^{24,40}, and the other calculated considering the percentage of vesicles in the sample^{22,23}. Both values plot within errors of our estimated CO_2/Nb ratios for undegassed basalts. Most MORB have been degassed; the extent of degassing can be calculated as the vertical distance between each sample and the solid line. Grey dashed lines represent different extent of degassing (10–90%) of the MORB samples. CO_2 content of MORB samples was from J. E. Dixon (personal communication) and A.E.S., unpublished data; Nb data was obtained from the MORB database (<http://petdb.ldeo.columbia.edu/petdb>).

source abundances of the volatile elements. Although this comparison has been made previously for some of the volatiles^{15–20}, it has not been possible for CO₂. This compound shows a positive correlation with Nb for the Siqueiros data (Fig. 2a), with an intercept that is within error of the origin. This behaviour suggests that CO₂ and Nb have similar partition coefficients during mantle melting. Because Nb is one of the most highly incompatible elements during melting, CO₂ also must exhibit highly incompatible behaviour. Although such behaviour has been inferred^{22–26}, it has not previously been demonstrated. The highly incompatible behaviour of CO₂ suggests that there is no stable carbon-rich phase (such as graphite) left behind after partial melting of the asthenosphere.

The Siqueiros melt inclusions are similar in their major-element compositions to primitive normal MORB, and indicators of the extent of melting such as Na₈ (wt% Na₂O corrected for crystal fractionation to 8% MgO; ref. 31) are within the range observed for normal MORB, suggesting no unusual extent of depletion for the Siqueiros samples. Therefore, the slightly larger depletion of the most incompatible trace elements compared to those of normal MORB cannot be the product of melting an extremely depleted MORB source, but rather represent (1) a slightly higher extent of melting from a normal asthenospheric source²⁹, (2) a lower extent of

melting from a source slightly more depleted, or (3) a melt from the depleted upper mantle that did not aggregate or mix with melts originated from an enriched source. Moreover, the Siqueiros picrites are isotopically (Sr, Nd, Pb, Hf, Th) indistinguishable from a normal MORB source³². Thus, the constant ratios of volatile elements to refractory trace elements of similar incompatibility in the Siqueiros melts should be generally applicable to the MORB source.

The CO₂/Nb ratios in the Siqueiros melt inclusions have an average value of 239 ± 46 (Supplementary Information Fig. 7a). Because the Nb content of the MORB mantle source is constrained to 0.3 ± 0.05 p.p.m. (refs 33, 34), this ratio suggests a CO₂ mantle source abundance of 72 ± 19 p.p.m. We can investigate the generality of this number by comparing the data to other rare MORB samples that appear to be undegassed, called ‘popping rocks’. These rocks have such high gas contents that they ‘pop’ when recovered from the sea floor to the surface. They have similar CO₂/Nb ratios to the Siqueiros samples, but at far higher overall abundance levels (Fig. 2b). Popping rocks have been used to make very high estimates of CO₂ content in the upper mantle^{22–24}. However, the popping rocks are enriched in incompatible trace elements such as Nb, and the high CO₂ contents therefore result from their enriched compo-

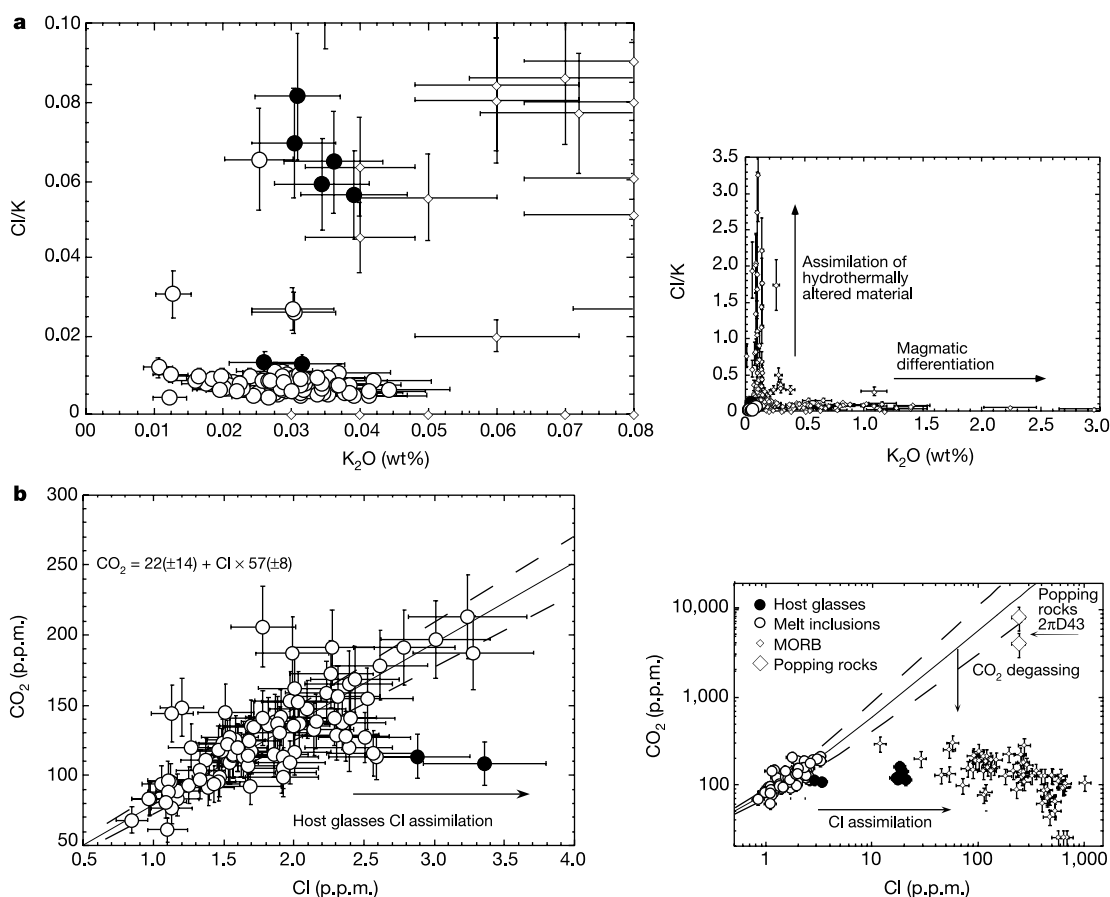


Figure 3 Evaluation of primary Cl content in Siqueiros and other MORB samples. **a**, Left panel; Cl/K ratios versus K₂O content for Siqueiros melt inclusions and host glasses compared to MORB. The MORB data are from ref. 20 and A.E.S., unpublished data. Right panel, the whole MORB data set²⁰ and the position of the Siqueiros samples. Two main trends can be distinguished in the right panel: one defined by assimilation of hydrothermally altered material, and the second defined by magmatic processes²⁰. **b**, Left panel; CO₂ contents versus Cl abundances in Siqueiros samples. Cl contents of the Siqueiros melt inclusions define a significant correlation with CO₂ (simple regression line, solid line). Right panel, the same data but using a logarithmic scale and including the values for other MORBs. The volatile undersaturation of the Siqueiros inclusions defines

the CO₂/Cl ratios for CO₂-undegassed and Cl-uncontaminated MORB (solid line, 2σ dashed line). In the Siqueiros inclusions, the CO₂ and Cl concentrations depend mainly on their behaviour during melting and their mantle source contents. Thus, Siqueiros melt inclusions have higher CO₂/Cl ratios than their host glasses and any other MORB samples, suggesting Cl contamination for the host glasses, and CO₂ degassing and/or Cl contamination for most MORBs. Popping rock (2IID43) plots within errors of our estimated CO₂/Cl ratios for undegassed/uncontaminated basalt. CO₂ and Cl contents of MORB samples are from J. E. Dixon (personal communication), and from C.H.L. and A.E.S. (unpublished data); part of the Cl data was obtained from the MORB database (<http://petdb.ldeo.columbia.edu/petdb>).

sitions. Although the popping rock CO_2 contents are not representative of normal MORB, the fact that these undegassed samples have the same CO_2/Nb ratio as the Siqueiros melt inclusions supports the mantle abundances and CO_2 systematics that we propose.

One of the characteristics of MORB is that the highly incompatible elements vary in concentration by two orders of magnitude. As CO_2 is also incompatible, it will have similar variations in primitive magmas that need to be taken into account in an evaluation of mantle source abundances and degassing histories. The CO_2/Nb ratio allows an estimate of the primary gas contents of any MORB with known Nb content, and therefore permits a first-order evaluation of the extent of degassing for any sample. Investigation of data from the literature in this context is shown in Fig. 2b. MORB data extend from CO_2/Nb ratios of about 239 to much lower values. The upper limit of previous data at our proposed mantle ratio is consistent with our reasoning, and shows the extensive degassing of CO_2 in most samples.

Volatile content of the depleted upper mantle

Previous estimates of the abundance of CO_2 in the mantle have ranged from 50 to $>2,000$ p.p.m. However, these estimates of CO_2 in the depleted upper mantle have used popping rocks^{22–24}, relied on mid-ocean-ridge CO_2 flux calculations with significant uncertainties⁹ or depended on general mantle degassing models^{8,25,26}. The value reported here is (to our knowledge) the first to make use of actual undegassed CO_2 abundances in normal MORB. Thus, we can estimate a carbon flux at ridges of $(9.3 \pm 2.8) \times 10^{11} \text{ mol yr}^{-1}$ using our CO_2 content of the upper mantle, a MORB flux³⁵ of $21 \text{ km}^3 \text{ yr}^{-1}$, and an assumption of a 10% extent of melting of the mantle source. This flux is within the range of some previous estimates^{8,26}, but is three times lower than the carbon flux calculated using popping rock (2IID43)^{22,23}. Likewise, we can provide improved estimates of the extent of ^3He degassing in MORB, the ^3He flux at ridges ($422 \pm 181 \text{ mol yr}^{-1}$), and the ^3He content of the upper mantle ($(7.5 \pm 3.2) \times 10^{-16} \text{ mol g}^{-1}$) using the relatively constant $\text{CO}_2/^3\text{He}$ ratio for MORB ($(2.2 \pm 0.7) \times 10^9$). These latter values are consistent with some of the most recent modelling estimates²⁶.

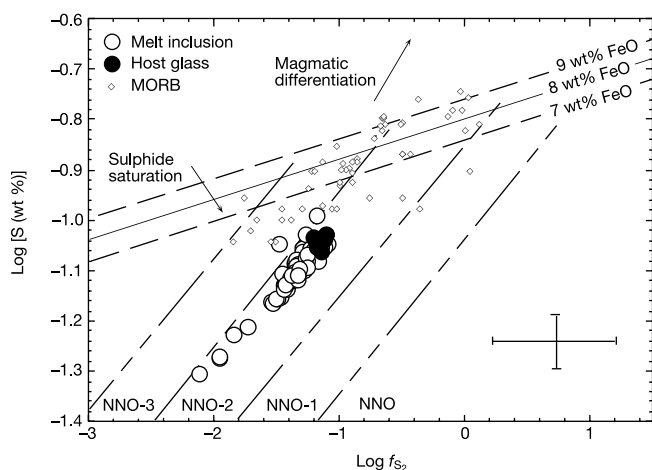


Figure 4 Relationship between sulphur fugacity f_{S_2} and dissolved sulphur in basaltic liquids along the sulphide saturation surface. The saturation surface (solid line) was calculated for the melt inclusion Siq 1-1 at $1,279^\circ\text{C}$ and modelled after ref. 17. Shown in the sulphide liquid undersaturated region are contours (short/long dashed lines) of constant oxygen fugacity, f_{O_2} , relative to Ni-NiO (NNO) buffer. Data for MORB are from ref. 17 and A.E.S., unpublished data. The dashed lines represent the field for sulphide saturation for a range in total FeO content from 7 to 9 wt%. Siqueiros melt inclusions range from within error of sulphide saturation to undersaturation. Error bars at lower right represent S analytical error (vertical) and model error (horizontal).

The undersaturated condition of the Siqueiros glasses also permits an evaluation of the mantle abundances of the other volatiles. The average $\text{H}_2\text{O}/\text{Ce}$ ratios in the melt inclusions (168 ± 95 (2σ); Supplementary Information Fig. 7b) are similar to other EPR glasses (averaging 168 ± 8 to 190 ± 41 ; refs 15, 16) and suggests a mantle abundance of 142 ± 85 p.p.m. (using an estimated upper-mantle Ce content of 0.84 ± 0.25 ; refs 33, 34), consistent with previous estimates of 100–190 p.p.m. for the Pacific MORB^{15,16,36}. The data confirm the similar geochemical behaviour of H_2O and Ce over a variable extent of melting proposed in ref. 15, as H_2O and Ce contents of our samples are lower than most EPR lavas.

Chlorine contents and Cl/K ratios for the Siqueiros melt inclusions (Cl mostly 1–3 p.p.m.; Cl/K = 0.0075 ± 0.0025 , Fig. 3a) reach the lowest values measured for MORB unaffected by hydrothermal alteration (Cl 2–400 p.p.m.; Cl/K ≈ 0.01 –0.08, range representing normal to enriched MORB, respectively²⁰). A few Siqueiros samples with higher Cl may have been slightly contaminated by sea water or altered crust, and are not included in these averages. The Siqueiros host glasses have higher Cl/K ratios (0.013–0.081) compared to the average Cl/K ratios in the melt inclusions (Fig. 3a). The increases in Cl/K in the host glasses, which have lower MgO contents than the melt inclusions, suggest that either a minor assimilation of hydrothermally altered material or interaction with sea water has occurred during crystallization and cooling of the lavas, consistent with the model of ref. 20. The low Cl contents of the melt inclusions show that Cl varies by two orders of magnitude or more in primitive MORB, and that Cl/K ratios can vary by an order of magnitude²⁰. This would make Cl among the most incompatible elements, similar to Ba or Rb in its behaviour as proposed in ref. 18 on the basis of considerations of ionic charge and radius. The Cl/K ratios from the Siqueiros samples place mantle Cl abundance for normal MORB at approximately 1 ± 0.5 p.p.m. (assuming the K content of the upper mantle is 100 ± 20 p.p.m.; refs 33, 34), which is a factor of five to ten lower than previous estimates for the source of normal MORB¹⁸. Moreover, because Cl and CO_2 contents in the Siqueiros melt inclusions define a significant correlation (Fig. 3b), using the estimated Nb content of the upper mantle of 0.3 ± 0.05 (refs 33, 34) and the CO_2/Nb ratios of 239 ± 48 , we obtain an upper-mantle Cl content of 0.9 ± 0.63 p.p.m. The calculated Cl content is almost identical to our estimate (Cl = 1 ± 0.5 p.p.m.) using the Cl/K ratios of the Siqueiros melt inclusions and the estimated K content of the upper mantle. The Cl and CO_2 correlation allows an estimate of the primary Cl content of any MORB with known Nb content, and therefore permits a first-order evaluation of the extent of Cl contamination for any basalt.

The average F content of normal MORB on the basis of limited data is approximately 250 ± 50 p.p.m. (ref. 18). The fluorine concentration in the Siqueiros samples is approximately half the normal values for MORB, averaging 85 p.p.m. for melt inclusions and 89 p.p.m. for host glasses (range from 50 to 135 p.p.m.). Fluorine has been proposed¹⁸ to behave similarly to phosphorus during magmatic processes, and an average F/P ratio for normal MORB of 0.4 ± 0.13 has been reported¹⁸. This average is within errors of the F/P ratios for Siqueiros host glasses and melt inclusions (average 0.27 ± 0.11 (2σ); see Supplementary Information Fig. 7c), and confirms the similar behaviour of F and P over a variable extent of melting. Therefore, using the estimated P content in the upper mantle of 54 p.p.m. (refs 32, 33), the Siqueiros F/P ratios constrain the F abundance in the normal MORB source to 16 ± 3 p.p.m., which is a factor of four lower than previous estimates for the source of normal MORB¹⁸.

The sulphur content of basaltic glasses is strongly dependent on temperature, oxygen fugacity, sulphur fugacity, bulk composition (primarily the FeO content of the lavas) and pressure¹⁹. Most primitive MORB are sulphide saturated, and contain about 800–1,000 p.p.m. S (ref. 17). The model of ref. 17 can be used to evaluate

whether the Siqueiros melt inclusions and host glasses are sulphide saturated. Calculations at 1,280 °C (the approximate liquidus temperature of the samples) show that although the host glasses are within error of the sulphide saturation curve, the melt inclusions range from barely saturated to sulphide-undersaturated (Fig. 4). S/Dy ratios in the Siqueiros melt inclusions vary by only 22% (2σ) with an average value of approximately 225 (see Supplementary Information Fig. 7d), demonstrating the similar behaviour of sulphur and heavy rare-earth elements. The difference in the degree of sulphide saturation between the host glasses and melt inclusions can be explained by crystal fractionation¹⁷. A sulphide-undersaturated magma will increase in sulphur content as the temperature decreases and fractional crystallization occurs, causing the basaltic melt to approach sulphide saturation. We estimate the sulphur abundance in the MORB mantle source to be 146 ± 35 p.p.m., using a S/Dy ratio of 225 ± 49 and a Dy content of 0.615 p.p.m. (refs 32, 33) in the upper mantle. These results show that any sulphide phase in the mantle source could have been nearly exhausted by the relatively large extents of melting involved in the production of the Siqueiros samples. Similar estimates of 190 ± 90 p.p.m. sulphur content for the MORB source have previously been made^{37–39}. These estimates relied on extrapolations of S abundances in variably melted and degassed mantle peridotitic xenoliths or metasomatized massif peridotites, and it is therefore not clear that they are representative³⁷. Our present estimate is consistent with the least altered and most fertile of these peridotites.

Our data provide improved estimates of the volatile content of the depleted upper mantle. These estimates are at the low end of what has been suggested on the basis of models published previously, and substantially lower than model estimates for the sources of hotspot basalts^{4–6}. The present data, therefore, support significant heterogeneity in the abundance of volatiles in the Earth's mantle. □

Received 21 March; accepted 16 August 2002; doi:10.1038/nature01073.

1. Gaetani, G. A. & Grove, T. L. The influence of water on melting of mantle peridotite. *Contrib. Mineral. Petrol.* **131**, 323–346 (1998).
2. Hirth, G. & Kohlstedt, D. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
3. Jung, H. Y. & Karato, S. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
4. Hauri, E. H. SIMS investigation of volatiles in volcanic glasses, 2: Abundances and isotopes in Hawaiian melt inclusions. *Chem. Geol.* (in the press).
5. Kingsley, R. H. & Schilling, J.-G. Carbon in Mid-Atlantic Ridge basalt glasses from 28°N to 63°N: Evidence for a carbon-enriched Azores mantle plume. *Earth Planet. Sci. Lett.* **129**, 31–53 (1995).
6. Simons, K., Dixon, J. E., Schilling, J.-G., Kingsley, R. & Poreda, R. Volatiles in basaltic glasses from the Easter-Salas y Gomez seamount chain and Easter microplate: implications for geochemical cycling of volatile elements. *Geochem. Geophys. Geosyst.* (in the press).
7. Javoy, M., Pineau, F. & Allègre, J. C. Carbon geodynamic cycle. *Nature* **300**, 171–173 (1982).
8. Zhang, Y. & Zindler, A. Distribution and evolution of carbon and nitrogen in Earth. *Earth Planet. Sci. Lett.* **117**, 331–345 (1993).
9. Moore, J. G. Water content of basalts erupted on the ocean floor. *Contrib. Mineral. Petrol.* **28**, 272–279 (1970).
10. Moore, J. G. Vesicularity and CO₂ in mid-ocean ridge basalt. *Nature* **282**, 250–253 (1979).
11. Bottinga, Y. & Javoy, M. Mid-ocean ridge basalt degassing: bubble growth and ascent. *Chem. Geol.* **81**, 255–270 (1990).
12. Dixon, J. E., Stolper, E. M. & Holloway, J. R. An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part I: calibration and solubility models. *J. Petrol.* **36**, 1607–1631 (1995).
13. Dixon, J. E. & Stolper, E. M. An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part II: applications and degassing. *J. Petrol.* **36**, 1633–1646 (1995).
14. Dixon, J. E. Degassing of alkalic basalts. *Am. Mineral.* **82**, 368–378 (1997).
15. Michael, P. J. Regionally distinctive sources of depleted MORB: evidence from trace elements and H₂O. *Earth Planet. Sci. Lett.* **131**, 301–320 (1995).
16. Danyushevsky, L. V., Eggins, S. M., Fallon, T. J. & Christie, D. M. H₂O abundance in depleted to moderately enriched mid-ocean ridge magmas. Part I: Incompatible behaviour, implications for mantle storage, and origin of regional variations. *J. Petrol.* **41**, 1329–1364 (2000).
17. Wallace, P. & Carmichael, I. S. E. Sulfur in basaltic magmas. *Geochim. Cosmochim. Acta* **56**, 1863–1874 (1992).
18. Schilling, J. G., Bergeron, M. B. & Evans, R. Halogens in the mantle beneath the North Atlantic. *Phil. Trans. R. Soc. Lond.* **297**, 147–178 (1980).
19. Carroll, M. R. & Webster, J. D. *Volatiles in Magmas* Reviews in Mineralogy (eds Carroll, M. R. & Holloway, J. R.) Vol. 30 231–279 (Mineralogical Society of America, Book Crafters, Chelsea, MI, 1994).
20. Michael, P. J. & Cornell, W. C. Influence of spreading rate and magma supply on crystallization and assimilation beneath mid-ocean ridges: Evidence from chlorine and major element chemistry of mid-ocean ridge basalts. *J. Geophys. Res.* **103**, 18325–18356 (1998).
21. Hauri, E. H., Newman, S. & Dixon, J. E. SIMS investigations of volatiles in volcanic glasses, 1: calibration, sensitivity and comparison with FTIR. *Chem. Geol.* (in the press).
22. Gerlach, T. M. Comment on “Mid-ocean ridge popping rocks: implications for degassing at ridge crests” by P. Sarda and D. Graham. *Earth Planet. Sci. Lett.* **105**, 566–567 (1991).
23. Graham, D. & Sarda, P. Reply to comment by T. M. Gerlach on “Mid-ocean ridge popping rocks: implications for degassing at ridge crests”. *Earth Planet. Sci. Lett.* **105**, 568–573 (1991).
24. Javoy, M. & Pineau, F. The volatile record of a “popping” rock from the mid-Atlantic ridge at 14° N: Chemical and isotopic composition of gas trapped in the vesicles. *Earth Planet. Sci. Lett.* **107**, 598–611 (1991).
25. Marty, B. & Jambon, A. C/³He in volatile fluxes from the solid Earth: implications for carbon geodynamics. *Earth Planet. Sci. Lett.* **83**, 16–26 (1987).
26. Marty, B. & Tolstikhin, I. N. CO₂ fluxes from mid-ocean ridges, arcs and plumes. *Chem. Geol.* **145**, 233–248 (1998).
27. Sobolev, A. V. & Chaussidon, M. H₂O concentrations in primary melts from supra-subduction zones and mid-ocean ridges: Implications for the H₂O storage and recycling in the mantle. *Earth Planet. Sci. Lett.* **137**, 45–55 (1996).
28. Wallace, P. & Anderson, A. T. *Encyclopedia of Volcanoes* (eds Haraldur, S., Houghton, B. McNutt, S. Rymmer, H. & Stix, J.) 149–170 (Academic, San Diego, 2000).
29. Perfit, M. R. *et al.* Recent volcanism in the Siqueiros transform fault: picritic basalts and implications for MORB magma genesis. *Earth Planet. Sci. Lett.* **141**, 91–108 (1996).
30. Roedder, E. *Fluid Inclusions* Reviews in Mineralogy (ed. Ribbe, P. H.) Vol. 12 (Mineralogical Society of America, Book Crafters, Chelsea, MI, 1984).
31. Klein, E. M. & Langmuir, C. H. Global correlation of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
32. Sims, K. W. *et al.* Chemical and isotopic constraints on the generation and transport of magma beneath the East Pacific Rise. *Geochim. Cosmochim. Acta* (in the press).
33. Hofmann, A. W. Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
34. Donnelly, K. E. *The Genesis of E-MORB: Extensions and Limitations of the Hot Spot Model* Thesis, Columbia Univ. (2002).
35. Crisp, J. A. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20**, 177–211 (1984).
36. Michael, P. J. The concentration, behaviour and storage of H₂O in the suboceanic upper mantle: implications for mantle metasomatism. *Geochim. Cosmochim. Acta* **52**, 555–566 (1988).
37. Lorand, J. P. Are spinel lherzolite xenoliths representative of the abundance of sulfur in the upper mantle? *Geochim. Cosmochim. Acta* **54**, 1487–1492 (1990).
38. Morgan, J. W. Ultramafic xenoliths: clues to Earth's late accretionary history. *J. Geophys. Res.* **91**, 12375–12387 (1986).
39. Chaussidon, M., Albarède, F. & Sheppard, S. M. Sulphur isotope variation in the mantle from ion microprobe analyses of micro-sulphide inclusions. *Earth Planet. Sci. Lett.* **92**, 144–156 (1989).
40. Bougault, H. *et al.* Mantle heterogeneity from trace elements: MAR triple junction near 14 degrees N. *Earth Planet. Sci. Lett.* **88**, 27–36 (1988).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Wang for assistance with the ion probe; C. Hadidiacos and C. Mandeville for help with electron probe analyses; J. Dixon for discussions and access to unpublished MORB volatile data; K. Donnelly, K. Simon, J. Y. Su and P. Asimow for discussions; and W. White and P. Michael for comments and suggestions that considerably improved the manuscript. This work was supported by the US NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E.S. (e-mail: asaa@ldeo.columbia.edu).

Production and detection of cold antihydrogen atoms

M. Amoretti*, C. Amsler†, G. Bonomi‡§, A. Bouchta‡, P. Bowe||, C. Carraro*, C. L. Cesar¶, M. Charlton#, M. J. T. Collier#, M. Doser‡, V. Filippini☆, K. S. Fine‡, A. Fontana☆☆, M. C. Fujiwara††, R. Funakoshi††, P. Genova☆☆, J. S. Hangst||, R. S. Hayano††, M. H. Holzschneider‡, L. V. Jørgensen#, V. Lagomarsino*‡‡, R. Landua‡, D. Lindelöf†, E. Lodi Rizzini§, M. Macri*, N. Madsen†, G. Manuzio*‡‡, M. Marchesotti☆, P. Montagna☆☆, H. Pruys†, C. Regenfus†, P. Riedler‡, J. Rochet‡‡, A. Rotondi☆☆, G. Rouleau‡‡, G. Testera*, A. Variola*, T. L. Watson# & D. P. van der Werf#

* Istituto Nazionale di Fisica Nucleare, Sezione di Genova, and ‡‡ Dipartimento di Fisica, Università di Genova, 16146 Genova, Italy

† Physik-Institut, Zürich University, CH-8057 Zürich, Switzerland

‡ EP Division, CERN, CH-1211 Geneva 23, Switzerland

§ Dipartimento di Chimica e Fisica per l'Ingegneria e per i Materiali, Università di Brescia, 25123 Brescia, Italy

|| Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark

¶ Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21945-970, and Centro Federal de Educação Tecnológica do Ceará, Fortaleza 60040-531, Brazil

Department of Physics, University of Wales Swansea, Swansea SA2 8PP, UK

☆ Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, and ** Dipartimento di Fisica Nucleare e Teorica, Università di Pavia, 27100 Pavia, Italy

†† Department of Physics, University of Tokyo, Tokyo 113-0033, Japan

A theoretical underpinning of the standard model of fundamental particles and interactions is CPT invariance, which requires that the laws of physics be invariant under the combined discrete operations of charge conjugation, parity and time reversal. Antimatter, the existence of which was predicted by Dirac, can be used to test the CPT theorem—experimental investigations involving comparisons of particles with antiparticles are numerous¹. Cold atoms and anti-atoms, such as hydrogen and antihydrogen, could form the basis of a new precise test, as CPT invariance implies that they must have the same spectrum. Observations of antihydrogen in small quantities and at high energies have been reported at the European Organization for Nuclear Research (CERN)² and at Fermilab³, but these experiments were not suited to precision comparison measurements. Here we demonstrate the production of antihydrogen atoms at very low energy by mixing trapped antiprotons and positrons in a cryogenic environment. The neutral anti-atoms have been detected directly when they escape the trap and annihilate, producing a characteristic signature in an imaging particle detector.

The current experiment, called ATHENA, seeks to compare the frequency of the 1S–2S electronic transition (ground state to first excited state) in hydrogen with that in antihydrogen. This frequency has been measured in hydrogen⁴ to an accuracy of 1.8 parts in 10^{14} . Obtaining similar precision in antihydrogen is possible in principle, but only if very cold (of the order of a few kelvin) anti-atoms are available.

Located adjacent to the antiproton decelerator⁵ (AD) ring at the CERN laboratory in Geneva, the ATHENA apparatus comprises four main subsystems: the antiproton catching trap, the positron accumulator, the antiproton/positron mixing trap, and the antihydrogen annihilation detector. All traps in the experiment are variations on the Penning trap⁶, which uses an axial magnetic field to transversely confine the charged particles, and a series of hollow cylindrical electrodes to trap them axially (Fig. 1a). The catching and mixing traps are adjacent to each other, and coaxial with a 3 T magnetic field from a superconducting solenoid. The

positron accumulator has its own magnetic system, also a solenoid, of 0.14 T. A separate cryogenic heat exchanger in the bore of the superconducting magnet cools the catching and mixing traps to about 15 K. The ATHENA apparatus⁷ features an open, modular design that allows great experimental flexibility, particularly in introducing large numbers of positrons into the apparatus—an essential factor in the current work.

The catching trap⁸ slows, traps, cools and accumulates antiprotons. To cool antiprotons, the catching trap is first loaded with 3×10^8 electrons, which cool by synchrotron radiation in the 3 T magnetic field. Typically, the AD delivers 2×10^7 antiprotons having kinetic energy 5.3 MeV and a pulse duration of 200 ns to the experiment at 100-s intervals. The antiprotons are slowed in a thin foil and trapped using a pulsed electric field. The antiprotons lose energy and equilibrate with the cold electrons by Coulomb interaction. The electrons are ejected before mixing the antiprotons with positrons. Each AD shot results in about 3×10^3 cold antiprotons for interaction experiments.

The positron accumulator, based on a design detailed in ref. 9, slows, traps and accumulates positrons emitted from a radioactive source (1.4×10^9 Bq ^{22}Na). Accumulation for 300 s yields 1.5×10^8 positrons¹⁰, 50% of which are successfully transferred to the mixing trap where they cool by synchrotron radiation. The mixing trap has

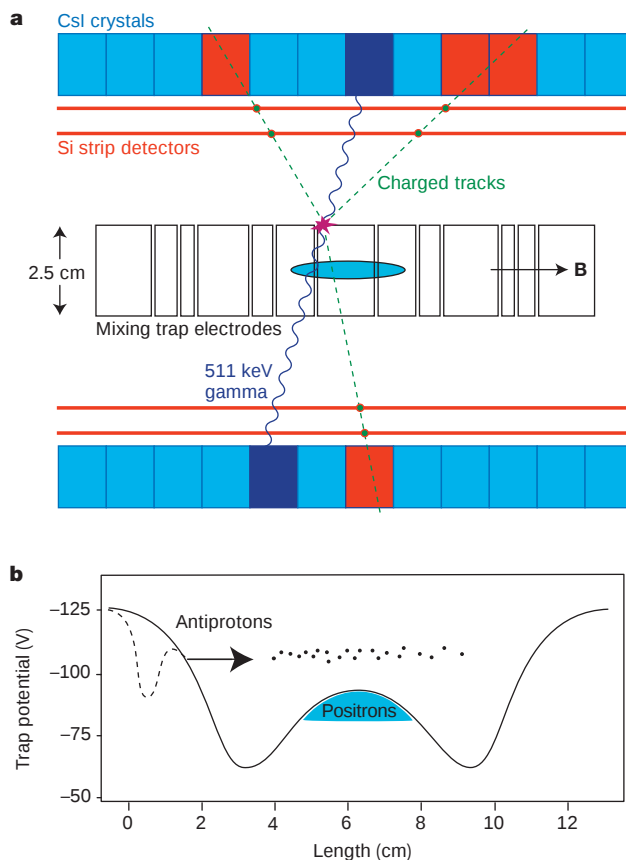


Figure 1 Central part of the ATHENA apparatus and trapping potential. **a**, Schematic diagram, in axial section, of the ATHENA mixing trap and antihydrogen detector. The cylindrical electrodes and the position of the positron cloud (blue ellipse) are shown. A typical antihydrogen annihilation into three charged pions and two back-to-back 511-keV photons is also shown. The arrow indicates the direction of the magnetic field. The detector active volume is 16 cm long and has inner and outer diameters of 7.5 cm and 14 cm, respectively. **b**, The trapping potential is plotted against length along the trap. The dashed line is the potential immediately before antiproton transfer. The solid line is the potential during mixing.

the axial potential configuration of a nested Penning trap¹¹ (Fig. 1b), which permits two plasmas of opposite charge to come into contact. In ATHENA, the spheroidal cloud of positrons can be characterized by exciting and detecting axial plasma oscillations, as demonstrated by ref. 12. Typical conditions are: 7×10^7 stored positrons, a radius of 2–2.5 mm, a length of 32 mm, and a maximum density of $2.5 \times 10^8 \text{ cm}^{-3}$.

Central to the observations reported here is the antihydrogen annihilation detector¹³ (Fig. 1a), situated coaxially with the mixing region, between the outer radius of the trap and the magnet bore. The detector is designed to provide unambiguous evidence for antihydrogen production by detecting the temporally and spatially coincident annihilations of the antiproton and positron when a neutral antihydrogen atom escapes the electromagnetic trap and strikes the trap electrodes. An antiproton typically annihilates into a few charged or neutral pions¹⁴. The charged pions are detected by two layers of double-sided, position-sensitive silicon microstrips. The path of a charged particle passing through both microstrip layers can be reconstructed, and two or more intersecting tracks allow determination of the position, or vertex, of the antiproton annihilation. The uncertainty in vertex determination is approximately 4 mm (1σ) and is dominated by the unmeasured curvature of the charged pions' trajectories in the magnetic field. The temporal

coincidence window is approximately 5 μs . The solid angle coverage of the interaction region is about 80% of 4π .

A positron annihilating with an electron yields two or three photons. The positron detector, comprising 16 rows, each row containing 12 scintillating, pure CsI crystals¹⁵, is designed to detect the two-photon events, consisting of two 511-keV photons that are always emitted back-to-back. The energy resolution of the detector is 18% full-width at half-maximum (FWHM) at 511 keV, and the photo-peak detection efficiency for single photons is about 20%.

The maximum readout rate of the whole detector is about 40 Hz. Ancillary detectors include large scintillator paddles external to the magnet, and a thin, position-sensitive silicon diode through which the incident antiproton beam passes before entering the catching trap.

To produce antihydrogen atoms, we fill a positron well in the mixing region with about 7×10^7 positrons and allow them to cool to the ambient temperature (15 K). The nested trap is then formed around the positron well. Next, approximately 10^4 antiprotons (three AD shots) are launched into the mixing region by pulsing the trap from one potential configuration (dashed line in Fig. 1b) to another (solid line). The mixing time is 190 s, after which all particles are dumped and the process repeated. Events triggering the imaging silicon detector (three sides hit in the outer layer) initiate readout of both the silicon and the CsI modules. We refer to this procedure as 'cold mixing'.

A second type of procedure, 'hot mixing', involves radio-frequency heating of the axial motion of the positron plasma. During the mixing cycle with antiprotons, the positrons are maintained at a temperature of several thousand kelvin. The positron temperature is monitored using the plasma mode analysis technique of ref. 12. The two mechanisms of antihydrogen formation expected to be important in ATHENA, radiative recombination and three-body recombination¹⁶, should be effectively suppressed at the hot mixing temperature, although the positrons and antiprotons still interact.

For comparison and for investigating background processes, we loaded antiprotons into the interaction region without positrons and accumulated detector data. The resulting data set is referred to as 'antiproton only'.

For cold mixing, interaction of the two particle species is immediately apparent, as the kinetic energy of the antiprotons is rapidly reduced (<1 s, an upper limit set by our current experimental method) to the level corresponding to the positron well. The cooling process is observed by lowering the wall of the antiproton well to various levels and recording the antiproton loss on the external scintillators. Such cooling has been reported previously¹⁷, but with a factor of 300 fewer positrons. Our measurements of antiprotons without positrons show no such cooling effect, also indicating effective removal of electrons.

At the start of cold mixing, we observe a rapid jump in the rate of antiproton annihilation, rising to a value that is tenfold higher than that for hot mixing. The positron loss during hot mixing is indistinguishable from that during cold mixing.

Subsequent observations are based on analysis of the annihilation detector data. Antiproton annihilation vertices are reconstructed from the silicon detector data using standard techniques¹⁸. For each vertex, we search for 'clean' evidence of 511-keV photons in the crystal data. A charged particle hit in a crystal, or an outer-layer silicon hit lying in the footprint of a crystal, excludes that particular crystal and its eight nearest neighbours. Next, we demand that exactly two of the remaining crystals have hits in an energy window around 511 keV, and that there are no hits of any energy adjacent to these two crystals. Energy calibration data, measured for each individual crystal, are used in this test.

To search for antihydrogen in the sample of events having a vertex and two 'clean' photons, we consider the opening angle, $\theta_{\gamma\gamma}$, between the lines connecting the vertex point to the geometric

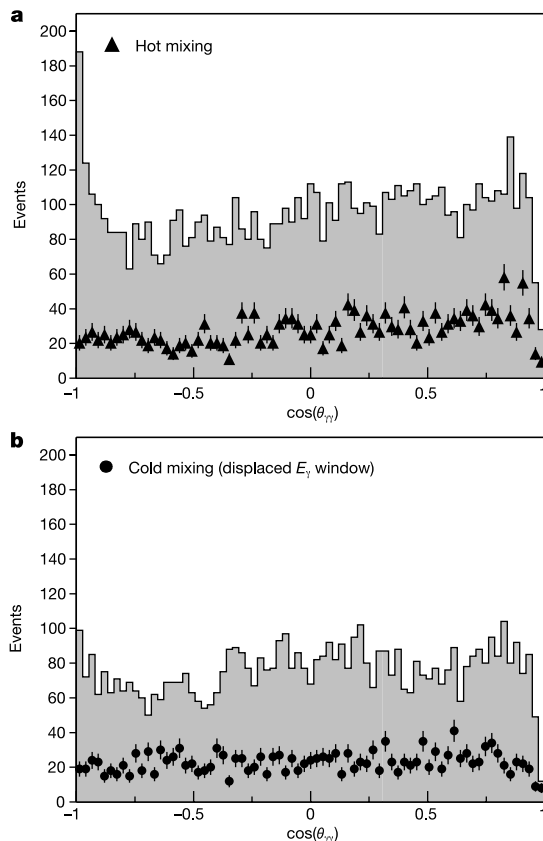


Figure 2 Experimental data. **a**, The number of events passing the selection criteria is plotted against the cosine of the opening angle $\theta_{\gamma\gamma}$ (see text for definition). The histogram is for cold mixing data (grey background). A total of 7,125 events, out of a sample of 103,270 reconstructed vertices, have two clean (but not necessarily back-to-back), detected photons in the 511-keV energy window. The data represent 165 mixing cycles. Filled triangles represent hot mixing data and are scaled by 1.6 to depict the same number of mixing cycles. **b**, The opening angle distribution (grey histogram) for antiproton-only data (99,610 vertices reconstructed, 5,658 clean events plotted). The filled circles represent cold mixing data, analysed using an energy (E_γ) window displaced upward so as not to include the 511-keV photo peak; no angular correlation of photons is seen.

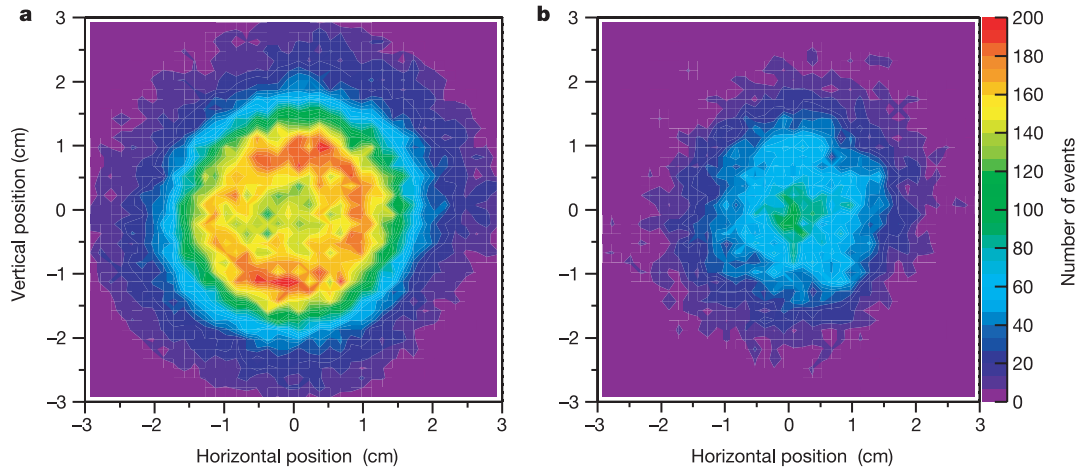


Figure 3 Colour contour plots of the distribution (obtained by projecting into the plane perpendicular to the magnetic field) of the vertex positions of reconstructed events. **a**, Cold mixing. All reconstructed antiproton annihilation vertices from the mixing region are plotted—no crystal cuts are applied. The trap inner radius is 1.25 cm. The

annihilations are centred on a slightly smaller radius, in agreement with our Monte Carlo simulations. (Some events appear to be outside of the trap radius owing to vertex reconstruction errors.) **b**, The same plot as above, but for hot mixing. These data are normalized to represent the same number of mixing cycles (165) as those in **a**.

centres of the two hit crystals. For an antihydrogen event, this angle should be 180° (or $\cos(\theta_{\gamma\gamma}) = -1$). The opening angle distribution for reconstructed events detected during cold mixing is shown in Fig. 2a. The equivalent plots for hot mixing (Fig. 2a) and antiproton-only (Fig. 2b) data are also shown. Compared with the hot mixing and antiproton-only data samples, the cold mixing data display a marked enhancement of events close to $\cos(\theta_{\gamma\gamma}) = -1$, indicating antihydrogen production. Analysis of the cold mixing data using an energy window that does not include the 511-keV peak in the crystal spectra yields no evidence for antihydrogen (Fig. 2b).

The shape of the $\cos(\theta_{\gamma\gamma})$ distribution predicted by Monte Carlo analysis of our detector for pure antihydrogen annihilation agrees well with the cold mixing measurement. In particular, the width of the peak close to $\cos(\theta_{\gamma\gamma}) = -1$ is the same for our data and for the Monte Carlo simulation. The dominant background for the experiment is caused by antiproton annihilations that contain a neutral pion, the decay of which produces electromagnetic showers in the material surrounding the trap and can lead to 511-keV signals in the CsI crystals. Our antiproton-only data (Fig. 2b) rule out this process as a source of antihydrogen-like signal in our cold mixing data. The shape of a Monte Carlo simulation of antiproton-only events also agrees well with our measurement. Other sources of background, such as cosmic rays and background from the accelerator complex (AD injection and extraction) have also been excluded. The probability for accidental temporal coincidence of antiproton and positron annihilation is negligible during mixing, as determined from measured trigger rates for each part of the detector. Furthermore, Monte Carlo simulations of this background indicate no enhancement in opening angle at 180° . Correlated loss due to scattering of antiprotons on positrons outside of the spheroidal cloud is ruled out by rate and dynamical considerations.

Further confirmation of antihydrogen production can be seen in the cold mixing data if the distribution of all antiproton annihilation vertex positions is projected onto a plane perpendicular to the magnetic field (Fig. 3a). A clear image of the trap electrodes is obtained, consistent with neutral anti-atoms annihilating on the wall. This observation also implies that many more antihydrogen atoms are produced than are actually detected through their angularly correlated photons, a fact consistent with our Monte Carlo analysis. The vertex distribution for mixing with heated positrons has a very different shape and no evidence of annihilation on the trap wall (Fig. 3b). Rather, the antiprotons annihilate in the

central region of the trap, probably owing to collisions with the residual gas or with trapped ions in the positron plasma. The average annihilation rate for hot mixing is about a factor of 4 smaller than that for cold mixing, whereas the peak ratio is about a factor of 10 smaller, as observed above.

If we consider cold mixing data only in the range $\cos(\theta_{\gamma\gamma}) < -0.95$, we have detected 131 ± 22 events with a reconstructed vertex and a pair of back-to-back, 511-keV photons above a conservatively scaled antiproton-only background. With an upper limit of the detection and reconstruction efficiency of 2.5×10^{-3} , on the basis of Monte Carlo analysis, we estimate that at least 50,000 antihydrogen atoms were created during cold mixing. This can be compared with the total number of antiprotons mixed, about 1.5×10^6 .

It is premature to discuss absolute production rates in our experiment, but it is noteworthy that the time distribution of the $\cos(\theta_{\gamma\gamma}) < -0.95$ events parallels the total annihilation rate during mixing. Further study is also needed to identify the mechanism of antihydrogen formation. We note that the present detection method does not identify the quantum state of the produced antihydrogen atoms. This is a question of considerable practical importance for future experiments involving trapping of neutral anti-atoms, as the demonstrated trapping techniques for hydrogen are only applicable to ground-state atoms. $\square$

Received 27 August; accepted 3 September 2002; doi:10.1038/nature01096.
Published online 18 September 2002.

1. Hagiwara, K. *et al.* The review of particle physics. *Phys. Rev. D* **66**, 010001 (2002).
2. Baur, G. *et al.* Production of antihydrogen. *Phys. Lett. B* **368**, 251–258 (1996).
3. Blanford, G. *et al.* Observation of atomic antihydrogen. *Phys. Rev. Lett.* **80**, 3037–3040 (1998).
4. Niering, M. *et al.* Measurement of the hydrogen 1S–2S transition frequency by phase coherent comparison with a microwave cesium fountain clock. *Phys. Rev. Lett.* **84**, 5496–5499 (2000).
5. Hellemans, A. Through the looking glass. *Nature* **406**, 556–558 (2000).
6. Dehmelt, H. Experiments with an isolated subatomic particle at rest. *Rev. Mod. Phys.* **62**, 525–530 (1990).
7. Fujiwara, M. C. *et al.* Producing slow antihydrogen for a test of CPT symmetry with ATHENA. *Hyperfine Interactions* (in the press).
8. Testera, G. *et al.* Toward the production of antihydrogen at rest in ATHENA. *Nucl. Instrum. Methods A* **461**, 253–255 (2001).
9. Surko, C. M., Gilbert, S. J. & Greaves, R. G. *Non-Neutral Plasma Physics* (eds Bollinger, J. J., Spencer, R. L. & Davidson, R. C.) **Vol. 3**, 3–12 (American Institute of Physics, New York, 1999).
10. Jorgensen, L. V., van der Werf, D. P., Watson, T. L., Charlton, M. & Collier, M. J. T. *Nonneutral Plasma Physics* (eds Anderegg, F., Schweikhard, L. & Driscoll, C. E.) **Vol. 4**, 35–44 (American Institute of Physics, New York, 2002).
11. Gabrielse, G., Rolston, S., Haarsma, L. & Kells, W. Antihydrogen production using trapped plasmas. *Phys. Lett. A* **129**, 38–42 (1988).
12. Tinkle, M. D., Greaves, R. G., Surko, C. M., Spencer, R. L. & Mason, G. W. Low-order modes as diagnostics of spheroidal non-neutral plasmas. *Phys. Rev. Lett.* **72**, 352–355 (1994).

13. Regenfus, C. A cryogenic silicon micro strip and pure-CsI detector for detection of antihydrogen annihilations. *Nucl. Instrum. Methods A* (in the press).
14. Bendiscioli, G. & Kharzeev, D. Antinucleon–nucleon and antinucleon–nucleus interaction, a review of experimental data. *Rivista Nuovo. Cim.* **17**(6), 1–42 (1994).
15. Amsler, C. *et al.* Temperature dependence of pure CsI: scintillation light yield and decay time. *Nucl. Instrum. Methods A* **480**, 494–500 (2002).
16. Charlton, M. & Humberston, J. W. *Positron Physics* (Cambridge Univ. Press, Cambridge, 2001).
17. Gabrielse, G. *et al.* First positron cooling of antiprotons. *Phys. Lett. B* **507**, 1–6 (2001).
18. Calligarich, E., Dolfini, R., Genoni, M. & Rotondi, A. A fast algorithm for vertex estimation. *Nucl. Instrum. Methods A* **311**, 151–155 (1992).

Acknowledgements

The authors comprise the ATHENA Collaboration. We would like to thank G. Bendiscioli, S. Bricola, P. Chiggiato, J. Hansen, H. Higaki, A. Lanza, C. Marciano, O. Meshkov, P. Salvini, G. Sobrero, B. Schmid and E. Sondergaard. We also thank the CERN AD team and C. Surko, who provided essential advice. This work was supported by Istituto Nazionale di Fisica Nucleare (Italy), Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ) e Fundação CCMN/UFRJ (Brazil), Grant-in-Aid for Creative Basic Research of Monbukagakusho (Japan), the Swiss National Science Foundation, the Danish Natural Science Research Council, The UK Engineering and Physical Sciences Research Council (EPSRC), The EU (Eurotraps Network), and the Royal Society. L.V.J., M.H.H., M.C. and J.S.H. acknowledge the work of the late B. Deutch.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to J.S.H.
(e-mail: jeffrey.hangst@cern.ch).

Quantum phase transition in a common metal

A. Yeh^{*†}, Yeong-Ah Soh^{*}, J. Brooke^{*†}, G. Aeppli^{*}, T. F. Rosenbaum[†]
& S. M. Hayden[‡]

^{*} NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

[†] The James Franck Institute and Department of Physics, The University of Chicago, Illinois 60637, USA

[‡] H.H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK

The classical theory of solids, based on the quantum mechanics of single electrons moving in periodic potentials, provides an excellent description of substances ranging from semiconducting silicon to superconducting aluminium. Over the last fifteen years, it has become increasingly clear that there are substances for which the conventional approach fails. Among these are certain rare earth compounds^{1,2} and transition metal oxides^{3,4}, including high-temperature superconductors^{5,6}. A common feature of these materials is complexity, in the sense that they have relatively large unit cells containing heterogeneous mixtures of atoms. Although many explanations have been put forward for their anomalous properties⁷, it is still possible that the classical theory might suffice. Here we show that a very common chromium alloy has some of the same peculiarities as the more exotic materials, including a quantum critical point⁸, a strongly temperature-dependent Hall resistance^{4,5} and evidence for a ‘pseudogap’⁹. This implies that complexity is not a prerequisite for unconventional behaviour. Moreover, it should simplify the general task of explaining anomalous properties because chromium is a relatively simple system in which to work out in quantitative detail the consequences of the conventional theory of solids.

Apart from complexity, a feature shared by almost all metals whose properties are at odds with conventional theory is proximity to magnetic instabilities^{10,11}, that is, the materials are nearly magnetic, and when placed under pressure or their chemical formulae

are slightly changed, they acquire magnetic order. A unique point in the associated pressure (or composition)–temperature phase diagrams is where the ordering temperatures vanish. Quantum rather than thermal fluctuations destroy the order beyond this point, which is therefore referred to as ‘quantum critical’.

We have selected chromium for our experiment because it is the simplest metal close to a quantum critical point. Indeed, it is a body-centred cubic (b.c.c.) antiferromagnet, whose magnetism¹² can be made to disappear with modest substitution of vanadium, its neighbour in the periodic table. On increasing the vanadium concentration x from 0, the Néel temperature T_N (Fig. 1a) and ordered magnetic moment¹³ (Fig. 1b) for $\text{Cr}_{1-x}\text{V}_x$ vanish continuously at the quantum critical point with $x_c = 0.035$. There is a wealth of information¹⁴ on this and other alloys of Cr, but, remarkably for such a simple substance, basic quantities such as the Hall resistance, which have engendered so much excitement for other materials from semiconductor heterostructures to high-temperature superconductors⁵, have not been studied systematically. The Hall effect has until now received little attention in the context of quantum phase transitions, and to our knowledge, our experiments are the first to examine the variation of the Hall effect near an antiferromagnetic quantum phase transition in Cr (ref. 15) or in any metal.

Our experiments were performed on single crystals, and the details are in the legend of Fig. 2. We plot in Fig. 2a the longitudinal resistivity $\rho(T)$ for several samples on each side of x_c . A key feature^{12,15} is the “chromium anomaly” at T_N , where the resistivity increases owing to a loss of the carriers localized through magnetic ordering. The anomaly rides on a background that consists of a composition-dependent constant and a T -dependent term. The former grows primarily with the disorder, which increases with increased (random) alloying, although there is a small anomaly at x_c (ref. 16). The latter does not vary significantly with composition, and, as demonstrated in the inset to Fig. 2a, appears to follow a

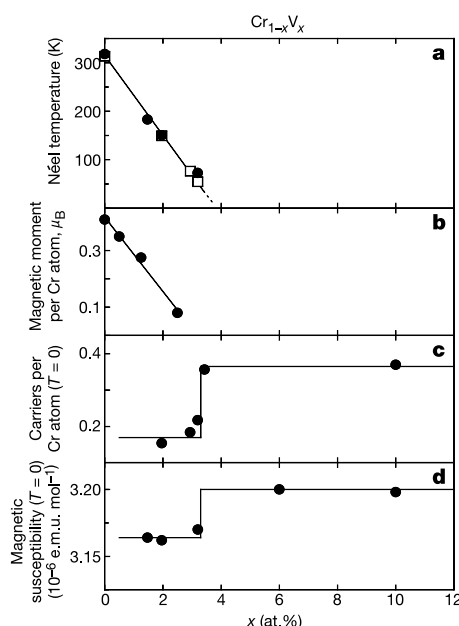


Figure 1 Magnetic ordering (Néel) temperature and zero-temperature properties of $\text{Cr}_{1-x}\text{V}_x$ near its quantum critical point. **a**, Phase diagram for the isostructural dilution series $\text{Cr}_{1-x}\text{V}_x$, where the filled circles and open squares denote Néel temperatures obtained from magnetic susceptibility and Hall effect data, respectively. **b**, Ordered magnetic moment measured by magnetic neutron diffraction¹³ at 4.2 K for $\text{Cr}_{1-x}\text{V}_x$. **c**, Low-temperature ($T < 5$ K) carrier density obtained from inverse of the Hall number. **d**, Low-temperature magnetic susceptibility.

simple T^3 power law at low T for all x . This is in sharp contrast to the rare-earth compound $\text{Ce}(\text{Cu},\text{Au})_6$ which has the most thoroughly studied^{1,2} quantum critical point among metals and shows $\rho(T)$ changing from T^2 a power-law to T and then back to T^2 again when passing through x_c . It also differs from various transition metal oxides where magnetic field or composition tuning achieves results^{6,17} similar to those for CeCu_6 . Furthermore, even the quadratic law $\rho(T) \propto T^2$, characteristic of conventional interacting Fermi liquids, is not detectable in $\text{Cr}_{1-x}\text{V}_x$ data. However, our experimental result agrees with one of the main predictions of the theory of antiferromagnetic quantum critical points in metals^{18,19}: the temperature dependence of the longitudinal resistivity remains indifferent to x .

The Hall effect, where the deflection of the carriers in a magnetic field gives rise to a voltage perpendicular to the current and field

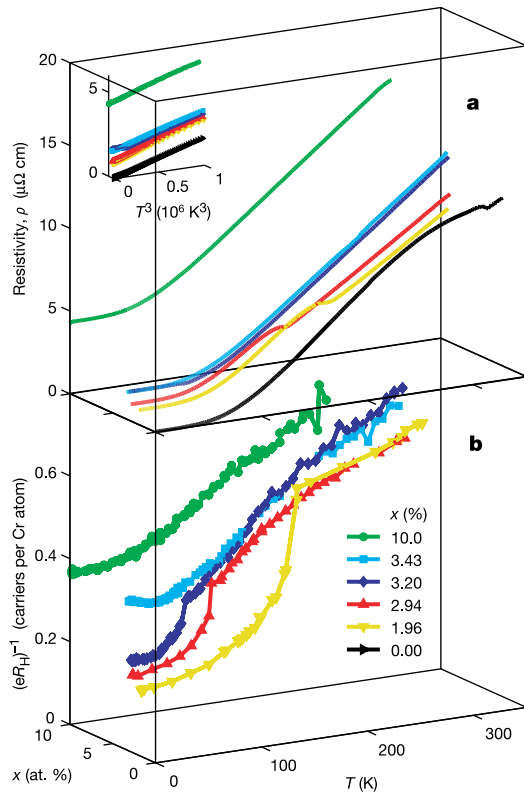


Figure 2 Temperature-dependent electrical properties near the quantum critical point of $\text{Cr}_{1-x}\text{V}_x$. Resistivity **(a)** and inverse Hall number **(b)** versus temperature T for various $\text{Cr}_{1-x}\text{V}_x$ samples. Inset in **a** displays the resistivity as a function of T^3 . Although the carrier density jumps at x_c , the T -dependent resistivity continues smoothly through the quantum critical point. The Hall number is positive for Cr and its alloys¹⁵, implying that holes are more important contributors than electrons. The samples used in our experiments are single crystals of $\text{Cr}_{1-x}\text{V}_x$ grown by Ames Laboratory in an arc zone refining furnace, usually from 99.995% pure Cr and 99.9% pure V and annealed for 72 h at 1,600 °C to relieve internal strain. Laue backscattering was used to determine the orientation of the b.c.c. lattice, and several pieces aligned with the (100) crystallographic direction were cut from the centre of each crystal using spark erosion. Proper etching to remove the surface oxide (with a 3:1::HCl:H₂O₂ solution quenched with THF) is essential for reproducible results. Increased temperatures (30 to 60 °C) and an ultrasonic bath were used to speed the etching process. For the resistivity and Hall measurements, the samples were parallelepipeds and small gold pads (~1,000 Å thick) were evaporated onto opposite sides of the parallelepipeds for use as electrical contacts. Indium solder and 1-mm-diameter gold wires connected the gold pads to the measurement circuit. The absolute value of the low-temperature (T) resistivity (~1 μΩ cm) indicates that the crystals are in the 'clean' limit, where disorder is unlikely to play a significant role. The magnetotransport measurements were performed in a top-loading ³He system using standard low-frequency lock-in methods.

directions, turns out to be far more sensitive to the quantum critical point. We plot in Fig. 2b the measured inverse Hall numbers, expressed as carrier densities, as a function of T for a variety of V concentrations. The spin-density-wave transition has a proportionately much larger effect on $1/R_H(T)$ than on $\rho(T)$; Fermi surface destruction manifests itself immediately as a change in the Hall number. The Hall number at low T increases by 100% in a narrow interval about x_c , jumping from one value weakly dependent on x for $x < x_c$ to an x -independent value for $x > x_c$ (Fig. 1c). The $T = 0$ phase transition seems to involve the destruction of a fixed amount of Fermi surface, and the quantum critical point simply occurs when this destruction is no longer possible. Even more remarkable is the strong temperature dependence of $1/R_H$ above $T_N(x)$, which is on the same scale as that which has attracted so much attention in the case of exotic metals such as the cuprates⁵. In the simplest interpretation, this effect implies an ever-increasing loss of carriers above the Néel temperature as x_c is approached from above.

The exceptional temperature dependence of the Hall coefficient in the high-temperature superconductors is accompanied by different temperature dependencies for the longitudinal and transverse conductivities⁵. We find similar behaviour for $\text{Cr}_{1-x}\text{V}_x$. We plot in Fig. 3a the ratio $\cot \theta_H$ of transverse to longitudinal conductivities at $H = 1.5$ T. The approximately T^2 dependence contrasts with the T^3 behaviour of the longitudinal resistivity (Fig. 2a), and is consistent with similar temperature dependencies seen not only for the high-temperature superconductors but also for V_{2-y}O_3 (ref. 4).

The Hall effect can be difficult to interpret because for multiple bands crossing the Fermi energy E_F it depends on carrier mobilities as well as densities²⁰, making it useful to examine another quantity sensitive to the carrier density at E_F . We have chosen to measure the bulk magnetic susceptibility χ , which for simple metals is proportional to the density of states at the Fermi level. Because of problems with surface oxides (see inset of Fig. 3b), measurements of

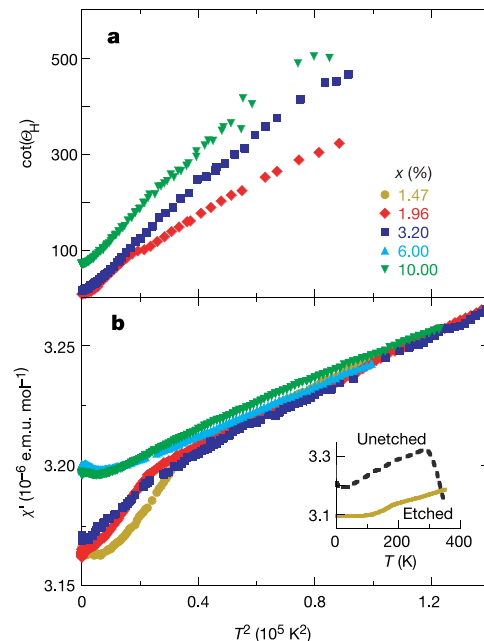


Figure 3 Electronic scattering and density of states, as measured by temperature dependence of electrical conductivity and magnetic susceptibility. **a**, Cotangent of Hall angle $\cot(\theta_H) = \sigma_{xx}/\sigma_{xy} = \rho/(HR_H)$, where σ is electrical conductivity and the field H is fixed at 1.5 T. **b**, Magnetic susceptibility χ' as a function of T^2 for various $\text{Cr}_{1-x}\text{V}_x$ samples. The data for χ' were all collected in fields of 3.2 and 3.6 T and cross-normalized to yield the same value above 370 K. Inset in **b** shows the importance of proper etch treatment of samples: the $x = 0.0147$ crystal used here does not reveal its intrinsic antiferromagnetic transition at 180 K unless the etch is performed.

the magnetic susceptibility of Cr and its alloys are rare, and if available, contradictory¹⁴. We have therefore collected our own data using a Quantum Design SQUID (superconducting quantum interference device) magnetometer. For ordinary insulating antiferromagnets, reducing T leads to a χ increasing towards a maximum at T_N , below which it decreases again. Chromium and its alloys behave quite differently (Fig. 3b): χ decreases in proportion to T^2 down to T_N , below which there is a sharp downturn which mirrors the behaviour of the inverse Hall coefficient (Fig. 2b). Here too, samples with $x < x_c$ and $x > x_c$ segregate into two classes (Fig. 1d), separated by a sharp change in the $T \rightarrow 0$ value of the susceptibility.

What is the origin of the increases of $1/R_H$ and χ with T in the paramagnetic regime on both sides of the quantum critical point? Ordinary band theory clearly yields the observed reduction of $1/R_H$ and χ on entering the ordered state for $x < x_c$, through a reduction in the density of states at the Fermi level arising because of the destruction of parts of the Fermi surface. Band theory can also account for increases of χ and $1/R_H$ with T in the paramagnetic regime with $T > T_N(x)$, provided that the Fermi energy lies at a minimum in the electronic density of states: warming would increase the density of states sampled. It is certainly possible for the minimum to be close to the Fermi energy, as has been computed for magnetically ordered Cr (ref. 21), although there are also calculations²² that show this not to be the case for paramagnetic Cr. Another prediction of band theory, which gives relatively small band masses for $\text{Cr}_{1-x}\text{V}_x$, is that as the content of V—with its outer electron count different from that of Cr—is varied, the Fermi level would shift significantly away from the minimum, flattening the T -dependence of $\chi(T)$ and $1/R_H(T)$ over the temperature ranges which we have examined. This is clearly not what we see in our data, which, in the paramagnetic phase, vary little with x , suggesting a minimum pinned at the Fermi level independently of band filling as regulated by V content.

The apparent failure of band theory in dealing with our transport data is also consistent with its failure (by a factor of 28) to account for the antiferromagnetic spin fluctuations^{23,24} in $\text{Cr}_{0.95}\text{V}_{0.05}$. This leaves us with the following facts to aid the search for an explanation of a minimum tied to the Fermi level: (1) $1/R_H$ decreases with T both in the paramagnetic phase, above $T_N(x)$, as well as upon entering the Néel state, below $T_N(x)$. As x increases towards x_c , the reduction in $1/R_H$ below T_N is diminished, while the amount by which $1/R_H$ decreases between 400 K and T_N is correspondingly increased. Taking $1/R_H$ as a measure of the effective Fermi surface area, there seems to be a trade-off between Fermi surface ‘lost’ below and above $T_N(x)$ —apparently the magnetic fluctuations found above $T_N(x)$ can be as powerful in diminishing the Fermi surface as is the magnetic order below $T_N(x)$. (2) An ordered spin-density wave induces a gap in the density of states for the Fermi surfaces involved. The gap is visible in optical data for Cr (ref. 25). (3) There are strong²⁴ antiferromagnetic fluctuations extending to very high energy (>0.4 eV) for paramagnetic $\text{Cr}_{1-x}\text{V}_x$. The fluctuations are patches of material with short-lived order of the type stabilized for infinite times and distances in the ordered phase.

Putting facts (1) to (3) together makes a pseudogap derived from the antiferromagnetic fluctuations the most probable cause of the minimum in the density of states seemingly pinned at the Fermi level. The idea is that over the lifetime of the antiferromagnetic critical fluctuations, a gap like that seen in the ordered phase (point (2)) will develop²⁶. Because the lifetime is not truly infinite at any finite temperature, we will see a smoothed and somewhat filled gap, or ‘pseudogap’, akin to what has caused so much excitement in the study of the copper oxides⁹. A direct measurement of the density of states, such as the magnetic susceptibility, will be sensitive to this pseudogap, as will transport properties involving timescales of the order of, or shorter than, the lifetime of the critical fluctuations. Both Figs 2b and 3b indicate that the pseudogap is felt deep into the

paramagnet: a substantive variation in $1/R_H(T)$ and $\chi(T)$ remains for V doping as high as 10%, three times x_c . That the variation is related to the magnetism of the undoped parent, rather than the consequence of conventional multiband effects, follows from point (1)—the variation induced by Néel order can be traded against that seen above $T_N(x)$. Furthermore, for $x = 0.05$, the inverse lifetime of the critical fluctuations is 0.1 eV (ref. 24), well below the width of the conduction band from which the carriers responsible for both metallicity and magnetism are drawn. This means that they live long enough to affect transport properties. Most notably, such fluctuations will tie down electrons from within the Fermi surface in the same way that magnetic order ties down electrons, implying an apparent loss of carriers on cooling as the fluctuations slow down. We conclude that in this common transition metal, we are finding pseudogap behaviour and a localization of charge, as evidenced by a highly temperature-dependent Hall effect, of the type hitherto associated only with exotic multicomponent materials such as the transition metal oxides.

We have performed the first measurements of the Hall effect—the simplest macroscopic probe of Fermi surface integrity—near a quantum critical point where metallic antiferromagnetism vanishes. The quantum critical point does not occur in a complex substance, but instead in a simple alloy whose properties should be calculable using a modern computer and the band theory of solids, thus providing an unusually rigorous baseline for finding unconventional behaviour. One aim of this paper is to encourage such computations of the Hall effect.

Our experiments yield two important results. The first, discussed in the previous paragraph, is that the T -dependent Hall effect is reminiscent of that for much more exotic metals, and, together with susceptibility data, implies a pseudogap in the spectrum of electronic excitations. The second is that the Hall effect evolves rather more quickly near the quantum phase transition than the magnetic order. Figure 1, which provides a summary of current knowledge about the quantum critical point in $\text{Cr}_{1-x}\text{V}_x$, puts the $T = 0$ charge and spin properties into perspective. The Hall effect (Fig. 1c), as well as our magnetic susceptibility data (Fig. 1d) indicate that as far as the Fermi surfaces are concerned, the $T = 0$ quantum phase transition is an either/or proposition. What this means is that as we scan vanadium composition x , we either see (for $x < x_c$) magnetic order involving a fixed number of carriers and a particular Fermi surface or (for $x > x_c$) disorder involving a different fixed number and another, larger Fermi surface. This quality contrasts with conventional expectations^{27,28} and the behaviour (Fig. 1b) of the order parameter itself, which vanishes more gradually as the quantum critical point is approached. Thus, near the quantum critical point, there are large amplitude fluctuations in the Fermi surface area, and these are likely to be connected to the anomalous finite temperature properties of $\text{Cr}_{1-x}\text{V}_x$, as has also been concluded² for the quantum critical point in $\text{Ce}(\text{Cu,Au})_6$. □

Received 17 July; accepted 5 August 2002; doi:10.1038/nature01044.

- Löhneysen, H. v. *et al.* Non-Fermi-liquid behavior in a heavy-fermion alloy at a magnetic instability. *Phys. Rev. Lett.* **72**, 3262–3265 (1994).
- Schröder, A. *et al.* Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407**, 351–355 (2000).
- Varma, C. M. *et al.* Phenomenology of the normal state of Cu-O high temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
- Rosenbaum, T. F., Hsueh, A., Carter, S. A. & Honig, J. M. Temperature dependence of the Hall angle in a correlated three-dimensional metal. *Phys. Rev. B* **57**, R13997–R13999 (1998).
- Chien, T. R., Brawner, D. A., Wang, Z. Z. & Ong, N. P. Unusual $1/T^3$ temperature dependence of the Hall conductivity in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. B* **43**, 6242–6245 (1991).
- Grigera, S. A. *et al.* Magnetic field-tuned quantum criticality in the metallic ruthenate $\text{Sr}_3\text{Ru}_2\text{O}_7$. *Science* **294**, 329–332 (2001).
- Allen, P. B. Superconductivity: Is kinky conventional? *Nature* **412**, 494–495 (2001).
- Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, Cambridge, 1999).
- Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an experimental overview. *Rep. Prog. Phys.* **62**, 61–122 (1999).
- Aeppli, G., Mason, T. E., Hayden, S. M., Mook, H. A. & Kulkarni, J. Nearly singular magnetic fluctuations in the normal state of a high- T_c cuprate superconductor. *Science* **278**, 1432–1435 (1997).
- Aeppli, G. & Broholm, C. *Handbook of the Physics and Chemistry of the Rare Earths* Vol. 19 123–175 (Elsevier, Amsterdam, 1994).

12. Fawcett, E. Spin-density-wave antiferromagnetism in chromium. *Rev. Mod. Phys.* **60**, 209–283 (1988).
13. Koehler, W. C., Moon, R. M., Trego, A. I. & Mackintosh, A. R. Antiferromagnetism in Chromium Alloys. I. Neutron Diffraction. *Phys. Rev.* **151**, 405–413 (1966).
14. Fawcett, E., Alberts, H. V., Galkin, V. Yu., Noakes, D. R. & Yakhmi, J. V. Spin-density wave antiferromagnetism in chromium alloys. *Rev. Mod. Phys.* **66**, 25–127 (1994).
15. de Vries, G. The transition in chromium and in some alloys of chromium with small amounts of other transition elements. *J. Phys. Rad.* **20**, 438–439 (1959).
16. Takeuchi, J., Sasakura, H. & Masuda, Y. Spin fluctuations in itinerant electron antiferromagnetic $\text{Cr}_{1-x}\text{V}_x$ system. *J. Phys. Soc. Jpn* **49**, 508–513 (1980).
17. Takagi, H. *et al.* Systematic evolution of temperature-dependent resistivity in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **69**, 2975–2978 (1992).
18. Mills, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
19. Hlubina, R. & Rice, T. M. Resistivity as a function of temperature for models with hot spots on the Fermi surface. *Phys. Rev. B* **52**, 9253–9260 (1995).
20. Furuya, Y. Temperature and magnetic field dependence of the Hall coefficient on the antiferromagnetic chromium. *J. Phys. Soc. Jpn* **40**, 490–497 (1976).
21. Hirai, K. Electronic structure of sinusoidal spin density wave state in chromium. *J. Phys. Soc. Jpn* **62**, 690–703 (1993).
22. Laurent, D. G., Callaway, J., Fry, J. L. & Brenner, N. E. Band structure, Fermi surface, Compton profile, and optical conductivity of paramagnetic chromium. *Phys. Rev. B* **23**, 4977–4987 (1981).
23. Staunton, J. B., Poulter, K., Ginatempo, B., Bruno, E. & Johnson, D. D. Incommensurate and commensurate antiferromagnetic spin fluctuations in Cr and Cr alloys from *ab initio* dynamical spin susceptibility calculations. *Phys. Rev. Lett.* **82**, 3340–3343 (1999).
24. Hayden, S. M., Doubble, R., Aeppli, G., Perring, T. G. & Fawcett, E. The strongly enhanced magnetic excitations near the quantum critical point of $\text{Cr}_{1-x}\text{V}_x$ and why strong exchange enhancement need not imply heavy fermion behavior. *Phys. Rev. Lett.* **84**, 999–1003 (2000).
25. Basov, D. N., Singley, E. J. & Dordevic, S. V. Sum rules and electrodynamics of high- T_c cuprates in the pseudogap state. *Phys. Rev. B* **65**, 054516 (2002).
26. Rosch, A. Some remarks on pseudogap behavior of nearly antiferromagnetic metals. *Phys. Rev. B* **64**, 174407 (2001).
27. Coleman, P., Pepin, C., Si, Q. & Ramazashvili, R. How do Fermi liquids get heavy and die? *J. Phys. Cond. Mat.* **13**, R723–R738 (2001).
28. Si, Q., Rabello, S., Ingersent, K. & Llewellyn Smith, J. Locally critical quantum phase transitions in strongly correlated metals. *Nature* **413**, 804–808 (2001).

Acknowledgements

We are grateful to P. Coleman and Q. Si for discussions. The work at the University of Chicago was supported by the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.A. (e-mail: gabe@research.nj.nec.com).

Light-induced conversion of an insulating refractory oxide into a persistent electronic conductor

Katsuro Hayashi*, Satoru Matsuishi†, Toshio Kamiya*†, Masahiro Hirano* & Hideo Hosono*†

* Transparent Electro-Active Materials Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Corporation, KSP C-1232, 3-2-1 Sakado, Tatatsu-ku, Kawasaki 213-0012, Japan

† Materials and Structures Laboratory, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8503, Japan

Materials that are good electrical conductors are not in general optically transparent, yet a combination of high conductivity and transparency is desirable for many emerging opto-electronic applications^{1–6}. To this end, various transparent oxides composed of transition or post-transition metals (such as indium tin oxide) are rendered electrically conducting by ion doping^{1–6}. But such an approach does not work for the abundant transparent oxides of the main-group metals. Here we demonstrate a process by which the transparent insulating oxide $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$ (refs 7–13) can be

converted into an electrical conductor. H^- ions are incorporated into the subnanometre-sized cages of the oxide by a thermal treatment in a hydrogen atmosphere; subsequent irradiation of the material with ultraviolet light results in a conductive state that persists after irradiation ceases. The photo-activated material exhibits moderate electrical conductivity ($\sim 0.3 \text{ S cm}^{-1}$) at room temperature, with visible light absorption losses of only one per cent for 200-nm-thick films. We suggest that this concept can be applied to other main-group metal oxides, for the direct optical writing of conducting wires in insulating transparent media and the formation of a high-density optical memory.

The crystal lattice of $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$ (refs 7–13) (C12A7) belongs to the space group of $I\bar{4}3d$ with a lattice constant of 1.199 nm (ref. 7), and the unit cell includes two molecules and twelve cages having a free space of $\sim 0.4 \text{ nm}$ in diameter (Fig. 1a). The chemical formula for the unit cell may be represented as $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} + 2\text{O}^{2-}$: the former denotes the lattice framework and the latter are called ‘free oxygen ions’. Thus each cage has a mean effective charge of $+1/3$ ($+4$ charge/12 cages), leading to the expectation that such a positively charged cage behaves like an F^+ centre^{14–17} upon capturing an electron. Another unique feature is the presence of the ‘free oxygen ion’, where an O^{2-} anion is accommodated in a cage to compensate for the positively charged framework^{7,8} and is coordinated with six Ca^{2+} cations that constitute a part of the cage wall. The inter-ionic spacing between the free oxygen ion and the Ca^{2+} cation is about 1.5 times longer than the sum (0.24 nm) of their ionic radii, indicating that the free oxygen ion is loosely bound in the cage. The cage has ‘entrances’ which are about 0.1 nm in diameter, and they may control mass transport between the inner cages and the outside. These features provide flexibility for replacing

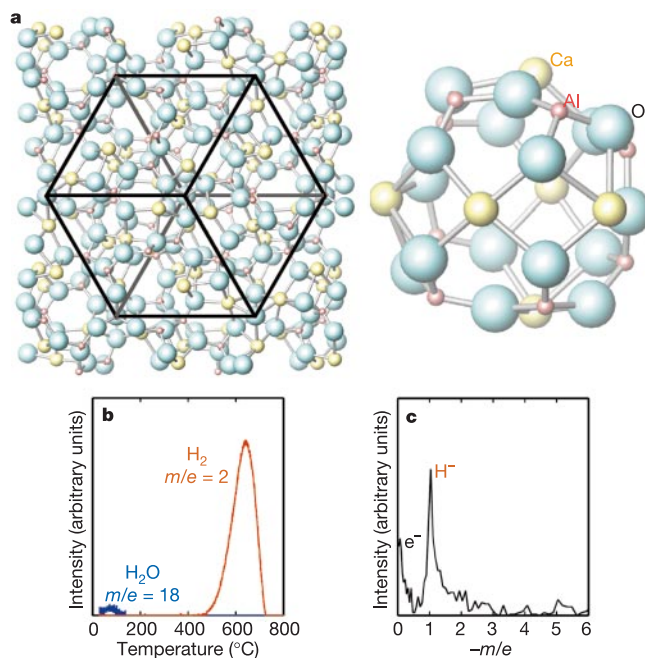


Figure 1 Incorporation of H^- ions in C12A7. **a**, Left is the crystal structure of C12A7 viewed from a $\langle 111 \rangle$ direction. A unit cell (lattice constant of 1.199 nm) with $Z = 2$ is indicated by a black frame. Right is the structure of a cage. Six Ca^{2+} ions coordinate to the centre of the cage. **b**, Thermogravimetric-mass spectroscopy (TG-MS) analysis on C12A7:H. Polycrystalline C12A7:H was heated at a rate of 10 K min^{-1} in a He atmosphere. Intense signal at $m/e = 2$ around 700°C is attributed to the H_2 molecule. **c**, Time-of-flight mass spectrum of field-extracted ions from polycrystalline C12A7:H. Extraction field is 300 V cm^{-1} . Sample temperature is 710°C . Two peaks at $-m/e = 0$ and 1 correspond to an electron and H^- , respectively.

the free oxygen ions with other anions such as OH^- (ref. 8), F^- , Cl^- (ref. 9), O_2^- (refs 10, 11) and O^- (ref. 11).

C12A7 polycrystalline samples were carefully fabricated to avoid extrinsic (impurity) effects using high-purity chemicals, CaCO_3 and Al_2O_3 . Single-crystals grown by a floating zone method¹² were used for most measurements to avoid the effects of grain boundaries and surfaces. Hydroxide ion (OH^-) concentration evaluated by infrared absorption spectroscopy was less than $2 \times 10^{19} \text{ cm}^{-3}$ in the as-grown crystal. The polycrystalline and single-crystal samples were heated at $1,300^\circ\text{C}$ for 2 h in a mixing gas composed of 20% H_2 /80% N_2 and was followed by rapidly cooling to room temperature. Secondary ion mass spectroscopy (SIMS) measurements for the single crystal revealed that thermal annealing increased the total hydrogenous species content to $2.5 \times 10^{20} \text{ cm}^{-3}$. The OH^- concentration increased slightly to around $5 \times 10^{19} \text{ cm}^{-3}$, which is only

about 20% of the total amount of hydrogenous species. $^2\text{H(D)}$ -nuclear magnetic resonance (NMR) measurements on the polycrystalline sample treated in 20% D_2 /80% N_2 atmosphere exhibited two peaks having comparable intensities at approximate chemical shifts of +20 and -20 p.p.m. The former is attributed to D^+ (OD^-) and the latter to D^- . This result verifies the presence of H^- (D^-) anions in the sample at a concentration comparable to that of OH^- (OD^-). The absence of a detectable signal in electron paramagnetic resonance (EPR) indicates the concentrations of paramagnetic ions or defect centres in the samples such as O_2^- (refs 10, 11), O^- (ref. 11), H_2^- ions and F^+ centre are negligibly small in the as-annealed state. When the polycrystalline samples were heated in a He carrier gas, hydrogen molecules were released between 500 and 750°C (Fig. 1b). In addition, applying an electric field¹³ at 700°C induced H^- ion emission (Fig. 1c). These results verify the incorporation of H^- anions^{18–20} in the sample through the following chemical reactions: (1) $\text{O}^{2-}(\text{cage}) + \text{H}_2(\text{atmosphere}) \rightarrow \text{OH}^-(\text{cage}) + \text{H}^-(\text{cage})$; or (2) $\text{O}^{2-}(\text{cage}) + \text{H}_2(\text{atmosphere}) \rightarrow 1/2\text{O}_2(\text{atmosphere}) + 2\text{H}^-(\text{cage})$.

The H^- -bearing samples (abbreviated as C12A7:H hereafter) look white in the polycrystalline powder or colourless and transparent for single crystals and are insulating (conductivity, σ , of $<10^{-10} \text{ S cm}^{-1}$ at room temperature). However, upon irradiating with ultraviolet light, two optical absorption bands centred at 2.8 and 0.4 eV are induced in the C12A7:H single crystal. Even after irradiation is stopped, these absorption bands remain unchanged and give rise to a persistent colour change from colourless to yellowish green (Fig. 2a). With an increase in the photon dose, the intensities of both bands were enhanced, keeping the relative intensity constant (Fig. 2b). These facts suggest that these two bands originate from an identical light-induced centre that will later be attributed to an F^+ -like centre created by trapping an electron in the cage.

The optical absorption edge of intrinsic C12A7 crystals is measured to be about 5 eV and exhibits a significant red-shift to about 4 eV for C12A7:H as seen in Fig. 2a. This suggests that the hydrogenous species controls the absorption edge in C12A7:H. Figure 2a also shows the relative efficiency of the colour centre yield as a function of the photon energy in the irradiated light. A fairly broad band peaks near 4.1 eV (Fig. 2a). This peak position is close to the absorption edge of the C12A7:H, indicating that the photo-excitation of the hydrogenous

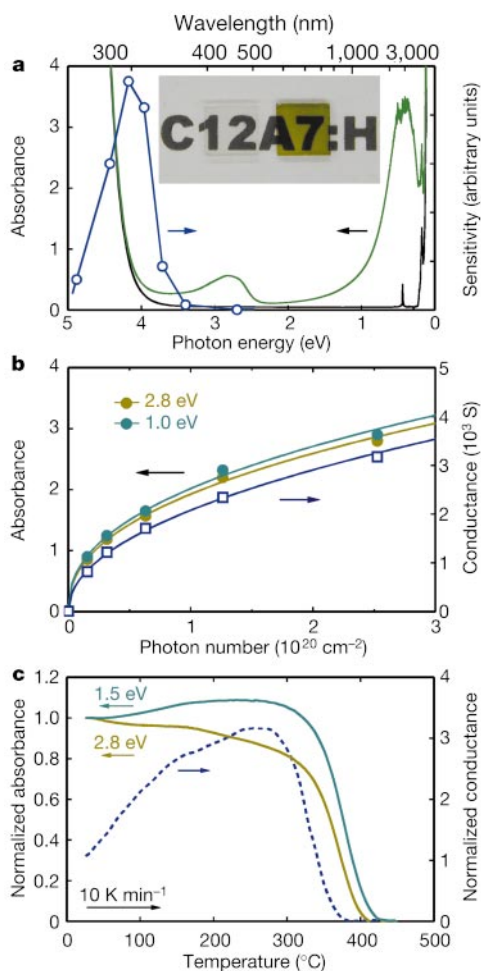


Figure 2 Insulator-conductor conversion of C12A7:H. **a**, The absorption spectra of C12A7:H single-crystals (0.3 mm thick) before (black) and after (green) irradiating with Xe-lamp light (4.9-eV light to $2 \times 10^{20} \text{ photons cm}^{-2}$). The absorbance change at 2.8 eV by irradiating $1.5 \times 10^{19} \text{ photons cm}^{-2}$ at each photon energy is plotted as blue circles, showing the sensitivity of the light-induced coloration. The inset is a photo of C12A7:H single crystals ($2.8 \times 2.8 \times 0.3 \text{ mm}^3$) before (left) and after (right) Xe-lamp light irradiation. **b**, The absorbance at 2.8 eV and 1.0 eV (as a measure of the 0.4-eV band because the absorbance of the 0.4-eV band became too strong to measure) and the conductance of a single crystal as a function of the irradiation flux of the 4.0-eV photon. **c**, The decay of coloration and conductance upon heating. Changes in the absorbance at 2.8 eV and 1.5 eV (as a measure of the 0.4-eV band because the *in situ* monitoring of the intensity in the near-infrared region is difficult owing to light emission from the heater that was used to raise the sample temperature) and conductance were recorded *in situ* during heating of 10 K min^{-1} in air. A C12A7:H single crystal was illuminated by Xe-lamp light for 15 min before heating. Each value was normalized by that at 25°C .

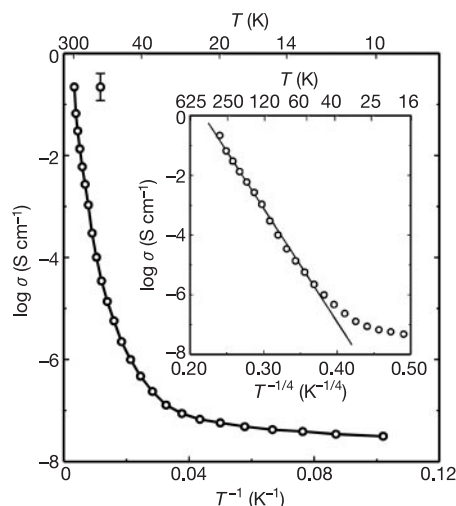


Figure 3 Electronic conduction in C12A7:H. Conductivity (σ)-temperature (T) characteristics of the ultraviolet-irradiated C12A7:H single crystal measured by impedance spectroscopy. The thickness of the light-induced coloration layer was determined by repeated mechanical thinning and optical absorption measurements and was estimated to be $\sim 200 \mu\text{m}$. Inset is a $\log \sigma$ versus $T^{-1/4}$ plot. The conduction may be controlled by a variable-range hopping of electrons.

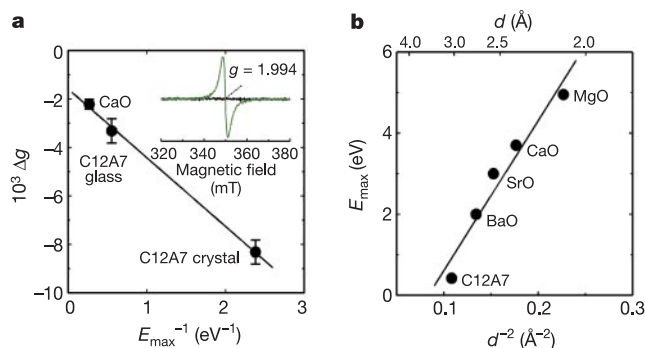


Figure 4 F^+ centre in C12A7:H. **a**, The relationship between the inverse of peak energy of absorption band ($1/E_{\max}$) and g-shift ($\Delta g = g - g_e$, where $g_e = 2.0023$) for the F^+ centre in CaO (ref. 14), C12A7 glass (ref. 17) and C12A7:H. The inset is the EPR spectra of polycrystalline C12A7:H before (black line) and after (green line) irradiation of Xe-lamp light. The X-band spectra were recorded at room temperature. **b**, The relationship between the peak energy ($E_{\max}$) and the inverse square of the cation-anion distance (d). Here, the sum of ionic radii is used as the d value for the alkaline-earth oxides and the average distance between the cage centre and the six coordinated Ca^{2+} ions is taken as that for C12A7:H.

species is responsible for this coloration.

Electrical conductivity of the single crystal was measured by complex impedance spectroscopy using an In-Ga alloy as electrodes. A drastic increase in the conductivity was observed along with the light-induced coloration. The σ value increased from $<10^{-10}$ to 0.3 S cm^{-1} at 300 K. The Seebeck coefficient of the conductive sample is negative ($-360 \mu\text{V K}^{-1}$), indicating n-type conduction. Coloration or conductivity changes were not observed for the samples before the H_2 treatment. The change in conductivity with the photon dose is similar to the absorbance change in the two light-induced bands (Fig. 2b). Figure 3 shows the temperature dependence of the conductivity in the light-irradiated C12A7:H single crystal. The σ value decreased from 0.3 to $10^{-7} \text{ S cm}^{-1}$ as the temperature was lowered from 300 to 10 K. Log (σ) is proportional not to T^{-1} but to $T^{-1/4}$ over a wide temperature range of 300 to 50 K. This characteristic indicates that the conduction is not controlled by a simple thermal activation process but more probably by a variable-range hopping²¹ of carriers. The light-induced conductive state remains unchanged even after illumination is stopped. But the conductivity rapidly dropped when the conductive sample was heated at $>320^\circ\text{C}$ (Fig. 2c). At this temperature the induced optical absorptions also rapidly decay (Fig. 2c), strongly suggesting that an identical species is responsible for both the two induced optical bands and the induced conductivity. When temperature rises above 550°C , H_2 gas is released from the sample (Fig. 1b) and the photosensitivity is lost. In other words, the emergence of the insulator-conductor conversion is reversible unless the heating temperature is high enough to release hydrogen. These observations indicate that the light-induced conduction is associated with the light-induced state changes of hydrogenous, probably H^- anion-related, species.

An EPR signal with an almost isotropic shape and a width (Δ_{msl}) of 2.4 mT appears at $g = 1.994$ as shown in the inset of Fig. 4a. Its intensity increases in proportion to those of the two induced optical absorption bands over irradiation time. The negative g-shift ($\Delta g = g - g_e$, where g_e is the value in a free space, 2.0023) indicates that the centre responsible for this EPR signal is of an electron-trapped type. This electron-trapped centre is coordinated by Ca^{2+} ions because the EPR signal does not show resolved hyperfine splitting (the natural abundance of nonmagnetic Ca is 99.9%). Figure 4a shows a linear relationship²² between $1/E_{\max}$ ($E_{\max}$ is the peak energy of an absorption band due to a F^+ centre) and Δg for

F^+ centres in a CaO crystal¹⁴, C12A7 glass (that is, $0.62\text{CaO}-0.38\text{Al}_2\text{O}_3$)¹⁷ and that of the EPR signal arising from a paramagnetic centre giving the 0.4 eV band in the C12A7:H crystal. This result attributes the EPR signal to an F^+ -like centre created in the cage inherent to C12A7.

Figure 4b plots $E_{\max}$ in alkaline-earth oxides¹⁶ together with that of the 0.4-eV band in the light-irradiated C12A7:H against the inverse square of the cation-anion distance (d^{-2}). The linear relationship between $E_{\max}$ and d^{-2} (the Mollow-Ivey rule¹⁵) for F^+ -centre in alkaline-earth oxides and the value in C12A7:H justifies the assignment of the 0.4-eV band to an F^+ -like centre in C12A7:H.

The maximum concentration of the F^+ -like centre determined by EPR was $2 \times 10^{19} \text{ cm}^{-3}$. The oscillator strength, f , of the 0.4-eV optical absorption band was about 0.3, which is reasonable for the $1s \rightarrow 2p$ allowed transition in F^+ centres (see for example, ref. 15). The f -value for the 2.8-eV band is an order of magnitude smaller than that of the 0.4-eV band. This band may be attributed to an allowed transition corresponding to $1s \rightarrow 2nd$ excited state in the same F^+ -like centre. We simulated the electronic structure in cages connected in series using a simple 'electron in a multiple potential well' model. It suggests that the energies of the light-induced optical absorption bands and the oscillator strengths are consistent with the wavefunction being spread over a few cages.

A possible mechanism for the photo-induced conversion of C12A7:H from an insulator to a conductor is proposed on the basis of these experimental findings. First, the ultraviolet irradiation induces an electron emission from the encaged H^- ions ($\text{H}^- \rightarrow \text{H}^0 + e^-$). Then, an empty cage electrostatically captures the photo-ionized electron and forms an F^+ -like centre. As the electron is weakly bound in the cage, the wavefunction would spread spatially. Consequently, electrons may migrate throughout the crystal by variable-range hopping. The resultant H^0 atoms recombine with each other to form H_2 molecules in a cage. The light-induced conduction is persistent below about 320°C , indicating that the recombination of the H_2 and the electron rarely occurs in such low temperatures.

We have proposed a concept for insulator-to-electronic conductor conversion using an extended wavefunction of F^+ centres and demonstrated the conductive state in a typical transparent main-group metal oxide (MGO), C12A7. F^+ centres are very common in oxides, including MGOs, so this concept may be applicable to other MGOs if the persistent F^+ centre concentration is large enough to produce conduction. A crystal composed of positively charged subnanometre-sized cages is appropriate.

As it is also possible to fabricate conductive (1 S cm^{-1}) C12A7 thin films with a transmission loss of less than 1% in the visible-light range, the present findings open up a new frontier in transparent oxide electronics. The persistent photo-induced conductive state allows for patterning conductive wires directly on transparent oxides by a photon mode, which will be essential in the emerging invisible-circuit technology⁶. These features also make the material attractive as rewritable recording media in multidimensional holographic memories. □

Received 21 May; accepted 7 August 2002; doi:10.1038/nature01053.

1. Ginley, D. S. & Bright, C. Transparent conducting oxides. *MRS Bull.* **25**, 15–21 (2000).
2. Hamberg, I. & Granqvist, C. G. Evaporated Sn-doped In_2O_3 films: Basic optical properties and applications to energy-efficient windows. *J. Appl. Phys.* **60**, R123–R160 (1986).
3. Robertson, J. Electronic structure of SnO_2 , GeO_2 , PbO_2 , TeO_2 and MgF_2 . *J. Phys. C* **12**, 4767–4776 (1979).
4. Kawazoe, H. et al. P-type electrical conduction in transparent thin films of CuAlO_2 . *Nature* **389**, 939–942 (1997).
5. Kawazoe, H., Yanagi, H., Ueda, K. & Hosono, H. Transparent p-type conducting oxides: Design and fabrication of p-n heterojunctions. *MRS Bull.* **25**, 28–37 (2000).
6. Thomas, G. Invisible circuit. *Nature* **389**, 907–908 (1997).
7. Bartl, H. & Scheller, T. Zur Struktur des $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$. *N. Jb. Miner. Mh.* **35**, 547–552 (1970).
8. Imlach, J. A., Glasser, L. S. D. & Glasser, F. P. Excess oxygen and the stability of " $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$ ". *Cement Conc. Res.* **1**, 57–61 (1971).

9. Jeevaratnam, J., Glasser, F. P. & Glasser, L. S. D. Anion substitution and structure of $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$. *J. Am. Ceram. Soc.* **47**, 105–106 (1964).
10. Hosono, H. & Abe, Y. Occurrence of superoxide radical ion in crystalline $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$ prepared via solid-state reaction. *Inorg. Chem.* **26**, 1192–1195 (1987).
11. Hayashi, K., Hirano, M., Matsuishi, S. & Hosono, H. Microporous crystal $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$ encaging abundant O^- radicals. *J. Am. Chem. Soc.* **124**, 738–739 (2002).
12. Watauchi, S., Tanaka, I., Hayashi, K., Hirano, M. & Hosono, H. Crystal growth of $\text{Ca}_{12}\text{Al}_{14}\text{O}_{33}$ by the floating zone method. *J. Cryst. Growth* **237**, 496–502 (2002).
13. Li, Q.-X. *et al.* Absolute emission current density of O^- from $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$. *Appl. Phys. Lett.* **80**, 4259–4261 (2002).
14. Henderson, B. & Wertz, J. E. Defects in the alkaline earth oxides. *Adv. Phys.* **17**, 749–855 (1968).
15. Agullo-Lopez, F., Catlow, C. R. A. & Townsend, P. D. *Point Defects in Materials* Ch. 5 (Academic, London, 1988).
16. Williams, R. T. & Friebele, E. J. *CRC Handbook of Laser Science and Technology* (ed. Weber, M. J.) Part I Vol. III (CRC, Boca Raton, FL, 1986).
17. Hosono, H., Asada, N. & Abe, Y. Properties and mechanism of photochromism in reduced calcium aluminate glasses. *J. Appl. Phys.* **67**, 2840–2847 (1990).
18. Giamello, E., Paganini, M. C., Murphy, D. M., Ferrari, A. M. & Pacchioni, G. A. Combined EPR and quantum chemical approach to the structure of surface F_2^+ (H) centres on MgO . *J. Phys. Chem.* **101**, 971–982 (1997).
19. Oliver, D., Hofmann, P. & Knözinger, E. H_2 chemisorption and consecutive uv stimulated surface reactions on nanostructured MgO . *Phys. Chem. Chem. Phys.* **1**, 713–721 (1999).
20. Hayward, M. A. *et al.* The hydride anion in an extended transition metal oxide array: $\text{LaSrCoO}_3\text{H}_{0.7}$. *Science* **295**, 1882–1884 (2002).
21. Mott, N. F. & Davis, E. A. *Electronic Processes in Non-crystalline Materials*, 2nd edn (Oxford Univ. Press, Oxford, 1979).
22. Weil, J. A., Bolton, J. R. & Wertz, J. E. *Electron Paramagnetic Resonance* (Wiley, 1994).

Acknowledgements

We thank I. Tanaka and S. Watauchi for growing the single crystals, S. Takeda for NMR measurements, and M. Sadakata, Q.-X. Li and T. Nishioka for TOF-MS measurements.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence or requests for materials should be addressed to K.H.
(e-mail: k-hayashi@net.ksp.or.jp).

Switch of flow direction in an Antarctic ice stream

H. Conway*, G. Catania*, C. F. Raymond*, A. M. Gades*, T. A. Scambos† & H. Engelhardt‡

* Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA

† National Snow and Ice Data Center, University of Colorado, Boulder, Colorado 80309, USA

‡ Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 99125, USA

Fast-flowing ice streams transport ice from the interior of West Antarctica to the ocean, and fluctuations in their activity control the mass balance of the ice sheet. The mass balance of the Ross Sea sector of the West Antarctic ice sheet is now positive—that is, it is growing—mainly because one of the ice streams (ice stream C) slowed down about 150 years ago¹. Here we present evidence from both surface measurements and remote sensing that demonstrates the highly dynamic nature of the Ross drainage system. We show that the flow in an area that once discharged into ice stream C has changed direction, now draining into the Whillans ice stream (formerly ice stream B). This switch in flow direction is a result of continuing thinning of the Whillans ice stream and recent thickening of ice stream C. Further abrupt reorganization of the activity and configuration of the ice streams over short timescales is to be expected in the future as the surface topography of the ice sheet responds to the combined effects of internal dynamics and long-term climate change. We suggest that

caution is needed when using observations of short-term mass changes to draw conclusions about the large-scale mass balance of the ice sheet.

Both the volume and areal extent of the West Antarctic ice sheet have decreased greatly since the early Holocene epoch^{2–4}, and there is debate about its future, partly because of its potentially large effect on global sea level^{5–7}. The grounding line in the Ross Sea sector has on average retreated at about 120 m yr^{-1} for at least 7,500 years; at this rate, complete deglaciation of the Ross Sea sector will take about 7,000 years (ref. 7). However, recent calculations based on synthetic aperture radar altimetry indicate overall growth of the Ross drainage system¹. It is unclear whether the present growth, which began less than two centuries ago, is due to unsteady flow of the ice streams, or if it signals the end of the Holocene retreat.

Fast-flowing ice streams dominate the modern drainage system of West Antarctica. Fast flow is enabled through lubrication of the bed by pressurized water and weak subglacial till^{8,9}, which allows rapid motion (10^2 to 10^3 m yr^{-1}) in spite of the very low surface slopes and associated low gravitational driving stress^{10,11}. The widths of the Ross ice streams are constrained by relatively stable ridges of slow-moving ice (Fig. 1), and spectacular zones of chaotic crevasses form¹² at the lateral margins of ice streams that are flowing faster than about 100 m yr^{-1} . Emerging evidence of former (now buried) shear margins and other relict flow features show that the configuration and activity of the Ross ice stream system has undergone major reorganization in the recent past; Siple ice stream stagnated 450 years ago (ref. 13; and B. E. Smith, unpublished results); the lower reaches of ice stream C (ISC) stagnated 150 years ago¹⁴; the Whillans ice stream (WIS) is now slowing and migrating^{15–17}. Such changes are indicative of a highly dynamic drainage system that is reminiscent of braided stream systems, which as a whole can be relatively stable even as individual channels exhibit rapid reorganization over short timescales^{18,19}.

Flow stripes and surface morphology similar to those seen on active ice streams^{11,15} are visible in the image of the eastern end of

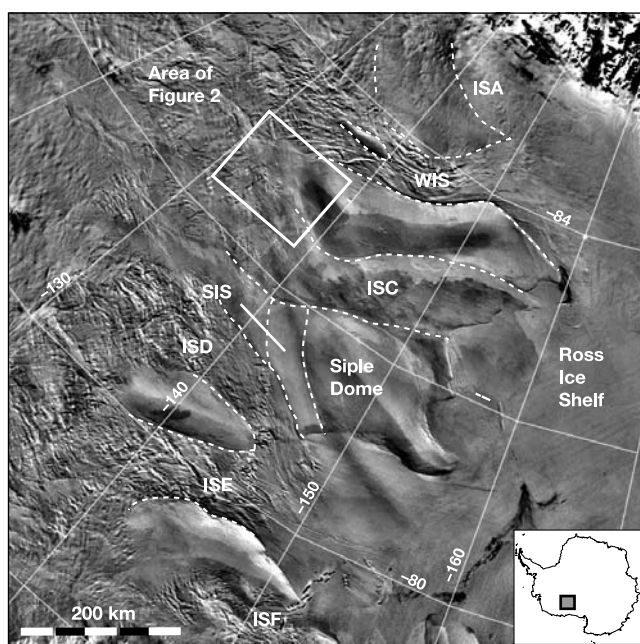


Figure 1 Satellite image of the Ross Sea ice drainage system in West Antarctica. White dashed lines delineate the lateral margins of present and former ice streams. The active ice streams (Whillans ice stream; WIS; ice streams A, D, E, F; ISA, ISD, ISE, ISF) flow at speeds up to 800 m yr^{-1} . Siple ice stream (SIS) stopped streaming 450 years ago (ref. 13; and B. E. Smith, unpublished results) and the lower reaches of ice stream C (ISC) stagnated about 150 years ago¹⁴.

Ridge BC (Fig. 2). Our ground-based radar profiles (Fig. 3) confirm the impression that ice once streamed across the ridge. The disturbed internal layering and diffractors, which are the tops of buried crevasses¹⁷, indicate that streaming flow (faster than about 100 m yr^{-1}) with associated high rates of marginal shearing has occurred in the past¹². We calculate the time of slowdown by tracking the deepest undisturbed radar layer across the top of the buried crevasses to a nearby core where the age–depth relationship has been measured²⁰. Results indicate that streaming flow ceased about 250 years ago, which is 100 years before the lower reaches of ISC stagnated, but 200 years after Siple ice stream stopped.

The elevation of the northern tributary of Whillans ice stream (WIS2) is now 120 m lower than ISC1 (a tributary of ISC), but the direction of past flow must have been toward the north into ISC for several reasons. (1) Analysis of the pattern of radar-detected internal layers across Ridge BC has shown that WIS2 has thinned about 200 m relative to ISC1 over the past 1,000 years (ref. 21). (2) The upper reaches of WIS2 are thinning²² and if the present rate (0.6 to 0.8 m yr^{-1}) has persisted since 250 years ago, it would have been 150–200 m higher then. ISC1 has thickened about 50 m since it stagnated²³, and so by this analysis WIS2 would have been 80–130 m higher than ISC1 250 years ago. Long-term thinning at half the present rate would still place WIS2 higher than ISC1 250 years ago. (3) The regional flow of ice is toward the northwest¹. Hence, we infer that this flow-path used to be a tributary to ISC, and because of its location to the south of tributary ISC1 (previously thought to be the

southern-most tributary of ISC) we name it ice stream C0 (ISC0).

Our measurements of surface velocity show that ice now flows southward out of ISC1 toward WIS2 (Fig. 2), which is almost completely opposite to the direction of past flow. The present-day flow, which cuts across the relict margin of ISC1 and other visible flow stripes (Fig. 2), follows closely the direction of surface slope. Observations suggest the following sequence of events: (1) about 1,000 years ago WIS2 started to thin faster than ISC for some reason that is unknown; (2) about 250 years ago thinning and inland migration of WIS2 beheaded tributary ISC0 and caused it to stagnate; (3) continuing thinning of WIS2 and more recent thickening of ISC1 has subsequently altered the surface topography in a way that has reversed the direction of both the gravitational driving

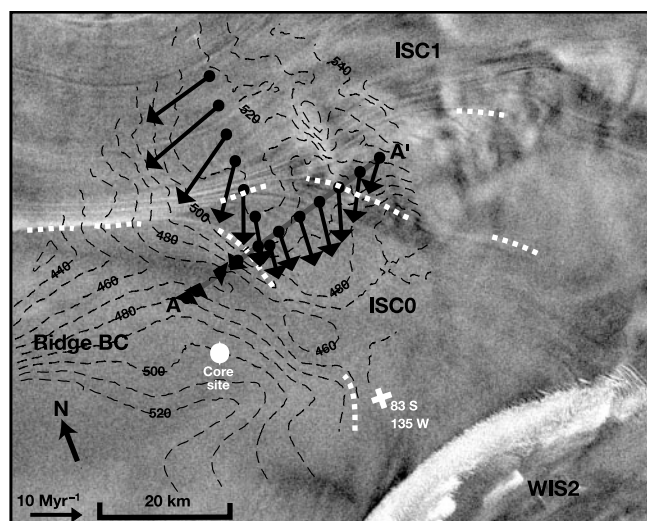


Figure 2 Flow features at the eastern end of Ridge BC (location shown in Fig. 1). The underlying image is derived from synthetic aperture radar data from the Radarsat Antarctic Mapping Mission (AMM-1) mosaic²⁸. Bright band at lower right is a result of scattering from open crevasses at the margin of the northern tributary of the Whillans ice stream (WIS2). Flow stripes, which are often visible on fast-flowing ice streams¹⁵, are also visible on the formerly active tributaries of ice stream C (ISC1 and ISC0). Black dashed lines, elevation contours (metres above WGS84) derived from GPS measurements and photoclinoetry²⁹. White dashed lines, relict ice stream margins inferred from buried crevasses that have been detected by others¹³ and by us using ice-penetrating radar. A–A', the location of the radar traverse shown in Fig. 3. The measured accumulation rate at the core site²⁰ is 8.8 cm yr^{-1} (ice equivalent). Crevasse tops are about 18 m below the surface along the southern margin of ISC1, which corresponds to an age of ~ 150 yrs before present, BP. Crevasses on the newly discovered tributary ISC0 are now buried 28 m below the surface (corresponding age is 250 yr BP). Present-day surface velocities (represented by arrows) were calculated from repeat GPS surveys (up to 13 months between surveys) of poles placed in the snow. The modern flow cuts across visible flow stripes; apparently the motion in the new direction has not yet been sufficient to deform the flow stripes significantly.

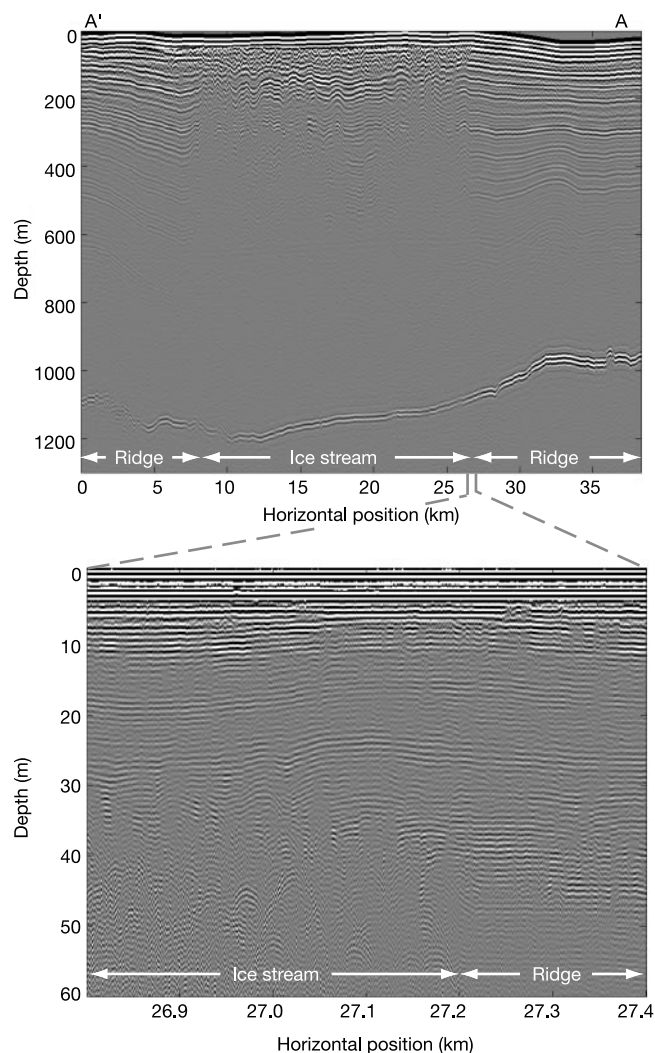


Figure 3 Ice-penetrating radar profiles along A–A' in Fig. 2. Present-day flow direction is into the page. Top, low-frequency (5 MHz) radar profile starts at $82.68^\circ \text{ S } 134.33^\circ \text{ W}$ and ends on Ridge BC. The upper 40 m are not resolved by the system. The bright reflector near 1,100 m is the bed. Internal layers on the ridges at both ends of the profile (visible to depths of 700 m) are continuous and relatively undisturbed. The disturbed internal layers between 6.5 km and 27.22 km are typical of those observed within active ice streams. Bottom, high-frequency (100 MHz) measurements show near-surface details of the transition across the western margin of the former ice stream. With this system the upper 5 m are not resolved, but outside the limits of the former ice stream (27.22 to 27.4 km), continuous internal layers are visible to depths greater than 60 m. Diffractors, visible between 26.8 and 27.2 km within the ice stream, are the tops of buried crevasses, indicating that high rates of strain have occurred in the past¹². The deepest undisturbed layer across the tops of the buried crevasses is 28 m below the surface.

stress and the flow of sub-glacial water (the basal hydraulic gradient is controlled primarily by the surface slope²⁴). This likely sequence of events raises the possibility that the shutdown of flow of ice and basal water through ISCO contributed to the eventual stagnation of the trunk region of ice stream C about 100 years later. Others have also suggested that the slow-down of ISC was a result of re-routing of ice²⁵ or basal water²⁶ away from the main trunk of the ice stream.

Of particular consequence are implications for the future. On the basis of the present-day geometry (ice thickness about 1.1 km, surface slope about 1.8×10^{-3}), measured surface velocities toward WIS2 (up to 20 m yr^{-1}) are two orders of magnitude faster than those expected from deformation of the ice column alone²⁷. Most of the speed today is a result of basal motion, which is not surprising because fast flow requires a lubricated bed, and liquid water has been detected in boreholes that have been drilled to the bed in the region⁸. Basal meltwater production will increase if flow continues to accelerate¹⁰, which could lead to streaming velocities in the near future. The mass balance of the Ross ice streams is now positive¹, but flow of 600 m yr^{-1} through such a hypothetical tributary (WIS3—1.1 km thick and 30 km wide) would discharge about $20 \times 10^{12} \text{ kg yr}^{-1}$ and effectively eliminate the present imbalance not only of ISC, but also of the entire Ross drainage system. □

Received 12 June; accepted 22 August 2002; doi:10.1038/nature01081.

- Joughin, I. & Tulacz, S. Positive mass balance of the Ross Ice Streams, West Antarctica. *Science* **295**, 476–480 (2002).
- Denton, G. H. & Hughes, T. J. Reconstructing the Antarctic Ice Sheet at the Last Glacial Maximum. *Quat. Sci. Rev.* **21**, 193–202 (2002).
- Shipp, S., Anderson, J. & Domack, E. Late Pleistocene–Holocene retreat of the West Antarctic ice-sheet system in the Ross Sea: Part 1 – Geophysical results. *Geol. Soc. Am. Bull.* **111**, 1486–1516 (1999).
- Conway, H., Hall, B. L., Denton, G. H., Gades, A. M. & Waddington, E. D. Past and future grounding-line retreat of the West Antarctic Ice Sheet. *Science* **286**, 280–283 (1999).
- Bentley, C. R. West Antarctic Ice Sheet collapse? *Science* **276**, 663–664 (1997).
- Oppenheimer, M. Global warming and the stability of the West Antarctic Ice Sheet. *Nature* **393**, 325–332 (1998).
- Bindschadler, R. A. Future of the West Antarctic Ice Sheet. *Science* **282**, 428–429 (1998).
- Kamb, B. in *The West Antarctic Ice Sheet: Behavior and Environment* (eds Alley, R. B. & Bindschadler, R. A.) Antarctic Research Series 77 157–199 (American Geophysical Union, Washington DC, 2001).
- Tulacz, S., Kamb, B. & Engelhardt, H. Basal mechanics of Ice Stream B, West Antarctica 2. Undrained plastic bed model. *J. Geophys. Res.* **105**, 483–494 (2000).
- Raymond, C. F. Energy balance of ice streams. *J. Glaciol.* **46**, 665–674 (2000).
- Whillans, I. M., Bentley, C. R. & van der Veen, C. J. in *The West Antarctic Ice Sheet: Behavior and Environment* (eds Alley, R. B. & Bindschadler, R. A.) Antarctic Research Series 77 257–281 (American Geophysical Union, Washington DC, 2001).
- Vaughan, D. G. Relating the occurrence of crevasses to surface strain rates. *J. Glaciol.* **39**, 255–266 (1993).
- Jacobel, R. W., Scambos, T. A., Nereson, N. A. & Raymond, C. F. Changes in the margin of Ice Stream C, Antarctica. *J. Glaciol.* **46**, 102–110 (2000).
- Retzlaff, R. & Bentley, C. R. Timing of stagnation of Ice Stream C, West Antarctica from short-pulse-radar studies of buried crevasses. *J. Glaciol.* **39**, 553–561 (1993).
- Fahnestock, M. A., Scambos, T. A., Bindschadler, R. A. & Kvaran, G. A millennium of variable ice-flow recorded by the Ross Ice Shelf, Antarctica. *J. Glaciol.* **46**, 652–664 (2000).
- Bindschadler, R. & Vornberger, P. Changes in the West Antarctic Ice Sheet since 1963 from declassified satellite photography. *Science* **279**, 689–692 (1998).
- Clarke, T. S., Liu, C., Lord, N. E. & Bentley, C. R. Evidence for a recently abandoned shear margin adjacent to Ice Stream B2, Antarctica, from ice-penetrating radar measurements. *J. Geophys. Res.* **105**, 13409–13422 (2000).
- Catania, G. & Paola, C. Braiding under glass. *Geology* **29**, 259–262 (2001).
- Murray, A. B. & Paola, C. A cellular model of braided rivers. *Nature* **371**, 54–57 (1994).
- Alley, R. B. & Bentley, C. R. Ice-core analysis on the Siple Coast of West Antarctica. *Ann. Glaciol.* **11**, 1–7 (1988).
- Nereson, N. A. & Raymond, C. F. The elevation history of ice streams and the spatial accumulation pattern along the Siple Coast of West Antarctica inferred from ground-based radar data from three inter-ice-stream ridges. *J. Glaciol.* **47**, 303–313 (2001).
- Shabtaie, S., Bentley, C. R., Bindschadler, R. A. & MacAyeal, D. R. Mass-balance studies of Ice Streams A, B and C, West Antarctica, and possible surging behavior of Ice Stream B. *Ann. Glaciol.* **11**, 165–172 (1988).
- Price, S. F., Bindschadler, R. A., Hulbe, C. L. & Joughin, I. R. Post-stagnation behavior in the upstream regions of Ice Stream C, West Antarctica. *J. Glaciol.* **47**, 283–294 (2001).
- Weertman, J. General theory of water flow at the base of a glacier or ice sheet. *Rev. Geophys. Space Phys.* **10**, 287–333 (1972).
- Rose, K. E. Characteristics of ice flow in Marie Byrd Land, Antarctica. *J. Glaciol.* **24**, 63–74 (1979).
- Anandakrishnan, S. & Alley, R. B. Stagnation of Ice Stream C, West Antarctica by water piracy. *Geophys. Res. Lett.* **24**, 265–268 (1997).
- Paterson, W. S. B. *The Physics of Glaciers* (Pergamon, Oxford, 1994).
- Jezek, K. C. Glaciological properties of the Antarctic ice sheet from RADARSAT-1 synthetic aperture imagery. *Ann. Glaciol.* **27**, 33–40 (1999).
- Scambos, T. A. & Fahnestock, M. A. Improving digital elevation models over ice sheets using AVHRR-based photogrammetry. *J. Glaciol.* **44**, 97–103 (1998).

Acknowledgements

We thank Raytheon Polar Services for logistical support in Antarctica, and C. R. Bentley, M. Conway, N. Lord and B. Smith for contributions. This work was supported by the US National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.C.

(e-mail: conway@ess.washington.edu).

Lateralization of magnetic compass orientation in a migratory bird

Wolfgang Wiltschko*, Joachim Traudt*, Onur Güntürkün†, Helmut Prior† & Roswitha Wiltschko*

* Zoologisches Institut, Fachbereich Biologie und Informatik, J.W. Goethe-Universität, Siesmayerstrasse 70, D-60054 Frankfurt am Main, Germany

† AE Biopsychologie, Fakultät für Psychologie, Ruhr-Universität Bochum, D-44780 Bochum, Germany

Lateralization of brain functions, once believed to be a human characteristic, has now been found to be widespread among vertebrates^{1–3}. In birds, asymmetries of visual functions are well studied, with each hemisphere being specialized for different tasks^{4–8}. Here we report lateralized functions of the birds' visual system associated with magnetoperception, resulting in an extreme asymmetry of sensing the direction of the magnetic field. We found that captive migrants tested in cages with the magnetic field as the only available orientation cue were well oriented in their appropriate migratory direction when using their right eye only, but failed to show a significant directional preference when using their left eye. This implies that magnetoreception for compass orientation, assumed to take place in the eyes alongside the visual processes^{9–11}, is strongly lateralized, with a marked dominance of the right eye/left brain hemisphere.

In birds, fibres of the optic nerves cross over completely and interhemispheric commissures are comparatively small. As a consequence, visual input from the right eye is predominantly processed by the left hemisphere and vice versa. Studies testing monocular birds with one eye occluded suggest a division of functions between the two hemispheres, with the left eye/right hemisphere being specialized for geometric aspects of visual cues and novelty⁸, whereas the right eye/left hemisphere predominantly processes object vision^{5,6,12}. These studies concerned tests performed in the small-scale surroundings of laboratories. However, recent studies with pigeons homing over distances of up to 40 km revealed that monocular birds using their right eye performed



Figure 1 A robin ready for monocular testing. Left, view of the covered eye; right, view of the open eye. J.T. took the photographs.

consistently better than those using their left eye, thus suggesting a superiority of the left brain hemisphere also in processes involving flight control, navigation and homing^{13,14}. The differences were mostly small, with one exception: in the only release conducted under overcast conditions, the difference in initial orientation and homing performance between the two groups was markedly increased¹³. This seemed to point out a possible involvement of the avian magnetic compass.

In birds, magnetic compass orientation is based on light-dependent processes assumed to take place in the eyes^{9–11}. Certain macromolecules are raised by photon absorption to singlet-excited states, forming radical pairs. By hyperfine coupling, singlet pairs are interconverted into triplet pairs. The triplet yield depends on the alignment of the molecules in the ambient magnetic field and could thus be used to detect directions⁹ (for details on the proposed mechanism, see ref. 9). Behavioural studies with passerines—using migratory orientation as a criterion of whether birds are able to obtain directional information from the magnetic field—showed that magnetoreception requires light from the blue–green part of the spectrum, with excellent orientation under near-monochromatic green light with a peak wavelength of 565 nm^{10,11,15}. Therefore, to test for a possible lateralization of the magnetic compass, we recorded the orientation behaviour of European robins, *Erithacus rubecula*, in cages under green light, testing the birds under binocular as well as under the two monocular conditions (left or right eye occluded).

For the monocular tests, one eye of the robin was covered with a cap so that light could reach the other eye only (see Fig. 1). Binocular and both types of monocular tests took place in the local geomagnetic field of Frankfurt am Main. The birds were additionally tested monocularly right-eyed in a magnetic field with the vertical component inverted, that is, inclination pointing upwards instead of downwards, to check whether they were using the magnetic inclination compass typical for birds^{15–17}.

We obtained two recordings of each bird in each of the four test conditions. The means of their two headings, α_b , together with the vector lengths based on the two headings, r_b , are given in Table 1. Under the binocular control condition, the birds significantly preferred their seasonally appropriate northerly migratory direction (Fig. 2a). The same is true when they were tested monocularly right-eyed in the local geomagnetic field (Fig. 2c); this distribution is statistically indistinguishable from control. In the monocular right-eyed condition, birds reversed their headings when the vertical component of the magnetic field was inverted (Fig. 2d), indicating the use of an inclination compass. There is no difference in variance

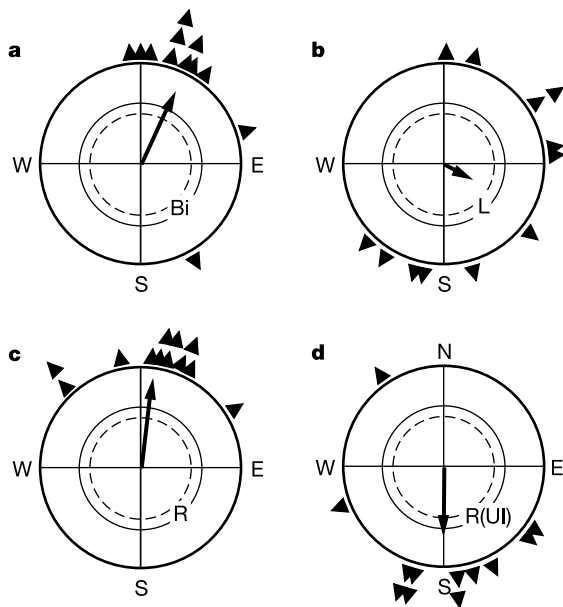


Figure 2 Orientation behaviour under monochromatic green light with the magnetic field as the only cue. The mean headings of the 12 birds are indicated as triangles at the periphery of the circle; the grand mean vector is represented by an arrow proportional to the radius of the circle (for numerical values, see Table 1). The inner circles are the 5% (dotted) and the 1% (solid) significance border of the Rayleigh test²⁵. **a**, Binocular control (Bi) tested in the geomagnetic field. **b**, Monocular left eye (L) tested in the geomagnetic field. **c**, **d**, Monocular right eye tested in the geomagnetic field (**c**; R) and in a magnetic field with the vertical component inverted, so that the inclination was pointing upwards (**d**; R(U)).

between this sample and the control sample or the right-eyed tests in the local geomagnetic field, although the preferred direction is significantly different ($P < 0.001$). In all three conditions, the mostly long vectors (r_b) indicate an excellent agreement between the headings of the two tests per bird (Table 1; see note on r_b in Methods). Together, these data suggest that being forced to use their right eye only did not adversely affect the birds' ability to orient with the help of their magnetic compass.

When the same birds were tested with their right eye occluded, however, the variance increased significantly. They failed to show a significant directional preference (Fig. 2b), with their mean head-

Table 1 Orientation performance of individual birds under test conditions

Bird	Binocular control (GF)		Monocular left eye (GF)		Monocular right eye (GF)		Monocular right eye (UI)	
	α_b	r_b	α_b	r_b	α_b	r_b	α_b	r_b
R1	72°	0.97	57°	0.39	9°	0.99	199°	0.98
R2	27°	0.94	213°	0.21	316°	0.84	200°	0.91
R3	4°	1.00	225°	0.66	58°	0.70	174°	1.00
R4	151°	0.91	16°	1.00	15°	0.94	325°	0.08
R5	17°	0.91	191°	0.89	21°	1.00	195°	0.92
R6	35°	0.97	165°	0.38	13°	0.96	130°	0.98
R7	16°	1.00	196°	0.47	5°	1.00	154°	0.98
R8	16°	1.00	56°	0.81	26°	0.98	196°	0.99
R9	359°	1.00	129°	0.60	23°	0.85	125°	0.99
R10	24°	0.89	81°	0.60	349°	0.81	164°	0.13
R11	25°	0.97	1°	0.60	318°	0.48	173°	0.92
R12	354°	0.84	85°	0.76	11°	0.82	249°	1.00
$n = 12$	25°	Median r_b	(119°)	Median r_b	7°	Median r_b	181°	Median r_b
	0.81***	0.97	0.31 (NS)	0.60	0.89***	0.90	0.70**	0.98

α_b and r_b are the direction and length of the mean vectors on the basis of the two recordings of the bird under the respective condition. The two bottom lines give the data of the second-order statistic: the direction and length of the grand mean vectors calculated on the basis of the directions of the 12 birds is given under α_b , with asterisks at the vector lengths indicating significance by the Rayleigh test (**, $P < 0.01$ and ***, $P < 0.001$; NS, not significant—here, the grand mean direction is given in parentheses; see also Fig. 2); the medians of r_b are given below. There is no statistical difference in the variance of α_b and in r_b between binocular control and the two monocular right eye conditions, whereas the α_b values of the monocular left eye condition are significantly more scattered ($P < 0.01$, Mann–Whitney U -test) and the r_b values are significantly shorter ($P < 0.01$, Wilcoxon test) than those recorded under control conditions. GF, geomagnetic field condition; UI, upward inclination condition.

ings scattered around the circle. (These headings seem to suggest a certain (nonsignificant) preference of a northeast–southwest axis. The distribution of the 24 headings on which the 12 mean headings are based, however, do not support bimodality, as the axial vector, 0.05, is even shorter than the unimodal vector, 0.19.) At the same time, the vector lengths (r_b) of the individual birds, with a median of 0.60, decreased significantly (see Table 1). These findings indicate that birds with monocular use of their left eye failed to orient, which implies that they were not able to derive the relevant directional information from the geomagnetic field.

Previous findings on the wavelength dependency of magneto-reception^{10,11} were always open to the argument that the observed disorientation under light of longer wavelengths was an unspecific response, as yellow and red light might have influenced the birds' motivation. Our present data make this type of explanation highly unlikely; they clearly indicate an involvement of the eye, in particular the right eye, in magnetoreception.

The performance of the birds when using only the right eye is indistinguishable from their binocular performance, whereas use of only the left eye obviously prevents normal magnetic orientation—this observation indicates a strong lateralization of the avian magnetic compass. It suggests that the processes leading to the perception of directional information from the magnetic field in birds occur almost exclusively in the right eye. As a consequence, the first steps of processing magnetic compass information must be expected to take place predominantly in the left hemisphere of the avian brain.

Such pronounced asymmetry may seem odd in animals whose organization is generally characterized by bilateral symmetry. However, the nature of the geomagnetic field as a stimulus does not require bilaterally symmetrical sensors, because, unlike in audition or vision, there is no need to survey large parts of the environment or to derive directional information from minute time differences between the left and right sensor. Birds could perceive the direction of the geomagnetic field with a non-paired sensory organ, provided it is built in an appropriate way; that is, shaped to include all axial directions⁹, a condition that could be satisfied by the hemispherical shape of one eye.

Electrophysiological evidence indicates that magnetic input is processed by neural structures of the accessory optic system, where single-unit responses to changes in the direction of the magnetic field have been recorded from the nucleus of the basal optic root in pigeons and passerine species. Similar responses have been recorded from the tectum opticum^{18–20}. Both of these midbrain structures feed into the nucleus rotundus^{21,22}, a thalamic relay activated after stimulation with changes in the direction of the ambient magnetic field²³. Tectum opticum, nucleus rotundus and the forebrain ectostriatum comprise the tectofugal system. In pigeons, lateralization of object vision depends on this tectofugal visual system²⁴, which displays numerous morphological asymmetries that are linked to visual lateralization at the behavioural level⁶. For instance, lateralization of object vision, in contrast to a number of other known lateralized functions, has a corresponding anatomical substrate. At the same time, its neuronal pathways are at least in part shared by magnetoreception for directional information. This raises an intriguing question about the origin of these neuronal asymmetries: did they emerge in connection with the one-sided system of magneto-reception in birds so that the superiority of the right eye/left hemisphere in object vision has been induced by already existing structures, or vice versa? Whatever the case, our findings provide evidence of lateralized reception of information for compass orientation, which is associated with a fairly strong hemispheric specialization. Given the basic need for orientation, this may open new perspectives for the understanding of why and how cerebral asymmetries have evolved. □

Methods

Test birds

European robins are small passerines that migrate at night. The test birds had been mist-netted in the garden of the Zoological Institute in Frankfurt am Main (50° 08' N, 8° 40' E) in September 2000; they were kept in an indoor bird room over the winter in a photoperiod simulating the local one until December, when it was reduced to a light/dark cycle of 8/16 h. After New Year, it was increased in two steps to a light/dark cycle of 13/11 h. This induced premature zugunruhe (migratory restlessness) so that testing took place between 8 January and 13 February 2001.

Test procedure

The robins were accustomed to wearing the eye covers—small spherical aluminium caps covering the eye, fixed to the bird's head with adhesive tape (Leukoplast; see Fig. 1)—by wearing them three times for about 90 min on each side before the tests. A pseudo-random sequence determined what bird was subjected to which treatment on a given day. Birds to be tested monocularly received their eye cap immediately before the test; it was removed again at the end of each test. Birds assigned as binocular controls received no treatment.

The tests began when the light went out in the birds' housing cages, and lasted approximately 75 min. Testing followed standard procedures^{10,11}. The birds were tested one at a time in funnel-shaped cages (35 cm upper diameter, 20 cm high) lined with typewriter correction paper; scratches were left in the coating as the birds moved. The cages were covered with an opaque plexiglas cover and placed in aluminium cylinders the top of which carried 24 green LEDs (peak frequency 565 nm, wavelength (λ)/2 at 553 and 583 nm, respectively). The light passed two diffusers before reaching the bird with an intensity of 2.1 mW m⁻².

Most tests took place in the local geomagnetic field of the Frankfurt am Main test site (46,000 nT, 66° inclination); the birds were also tested with their left eye covered in a magnetic field of equal intensity, but with the vertical component inverted (46,000 nT, -66° inclination) produced by a pair of Helmholtz coils.

Data analysis and statistics

The scratches were counted by experimenters that were blind to the test condition, and from the distribution of scratches the heading of the bird in that test was determined. One recording under control conditions with fewer than the standard limit of 35 scratches was excluded from the analysis because of too little migratory activity; this recording was repeated. From the two headings per test condition, the mean heading α_b and a vector length r_b of each bird were calculated (for details, see ref. 11), the latter indicating how well the headings of the two tests coincide. Note that the vector length, r , resulting from two data points, does not change linearly with increasing angular distance, $\Delta\alpha$, between the two points, but follows the function $r = \cos(\Delta\alpha/2)$. Therefore, long vectors are over-represented, with the median at 90° being 0.71 and the quartiles at 45° and 135° being 0.92 and 0.38, respectively.

The mean headings (α_b) of the twelve birds in each condition were combined to the grand mean vector, which was tested for directional preference using the Rayleigh test²⁵. The Watson Williams test was used to test for differences in direction and the Mann Whitney *U*-test was applied to the angular differences from the mean to test for differences in the variance²⁵. The vector r_b values were compared with those obtained under control conditions using the Wilcoxon test for matched pairs of data.

Received 21 February; accepted 21 June 2002; doi:10.1038/nature00958.

- Vallortigara, G. Comparative neuropsychology of the dual brain: a stroll through animals' left and right perceptual worlds. *Brain Lang.* **73**, 189–219 (2000).
- Rogers, L. & Andrew, R. J. (eds) *Comparative Vertebrate Lateralization* (Cambridge Univ. Press, Cambridge, 2002).
- Hunt, G. R., Corballis, M. C. & Gray, R. D. Laterality in tool manufacture by crows. *Nature* **414**, 707 (2001).
- Clayton, N. Lateralization and unilateral transfer of spatial memory in marsh tits. *J. Comp. Physiol. A* **171**, 799–806 (1993).
- Rogers, L. Behavioral, structural and neurochemical asymmetries in the avian brain: a model system for studying visual development and processing. *Neurosci. Biobehav. Rev.* **20**, 487–503 (1996).
- Güntürkün, O. Avian visual lateralization: a review. *NeuroReport* **8**, 3–11 (1997).
- Vallortigara, G., Regolin, L., Bortolomol, G. & Tommasi, L. Lateral asymmetries due to preferences in eye use during visual discrimination learning in chicks. *Behav. Brain Res.* **74**, 135–143 (1996).
- Tommasi, L. & Vallortigara, G. Encoding of geometric and landmark information in the left and right hemispheres of the avian brain. *Behav. Neurosci.* **115**, 602–613 (2001).
- Ritz, T., Adem, S. & Schulten, K. A model for vision-based magnetoreception in birds. *Biophys. J.* **78**, 707–718 (2000).
- Wiltchko, W., Munro, U., Ford, H. & Wiltchko, R. Red light disrupts magnetic orientation of migratory birds. *Nature* **364**, 525–527 (1993).
- Wiltchko, W. & Wiltchko, R. Light-dependent magnetoreception in birds: the behaviour of European robins, *Erithacus rubecula*, under monochromatic light of various wavelengths and intensities. *J. Exp. Biol.* **204**, 3295–3302 (2001).
- Prior, H. & Güntürkün, O. Parallel working memory for spatial location and food-related object cues in foraging pigeons: binocular and lateralised monocular performance. *Learn. Memory* **8**, 44–51 (2001).
- Ulrich, C. I. et al. Left-hemispheric superiority for visuospatial orientation in homing pigeons. *Behav. Brain Res.* **104**, 169–178 (1999).
- Prior, H., Wiltchko, R., Stapp, K., Güntürkün, O. & Wiltchko, W. *Orientation and Navigation—Birds, Humans and other Animals* 25–1–25–8 (Royal Institute of Navigation, Oxford, 2001).
- Wiltchko, W., Gesson, M. & Wiltchko, R. Magnetic compass orientation of European Robins under 565 nm green light. *Naturwissenschaften* **88**, 387–390 (2001).
- Wiltchko, W. & Wiltchko, R. Magnetic compass of European Robins. *Science* **176**, 42–64 (1972).
- Wiltchko, R. & Wiltchko, W. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).

18. Semm, P., Nohr, D., Demaine, C. & Wiltchko, W. Neural basis of the magnetic compass: interaction of visual, magnetic and vestibular inputs in the pigeon's brain. *J. Comp. Physiol. A* **155**, 283–288 (1984).
19. Semm, P. & Demaine, C. Neurophysiological properties of magnetic cells in the pigeon's visual system. *J. Comp. Physiol. A* **159**, 619–625 (1986).
20. Beason, R. C. & Semm, P. Magnetic responses of the trigeminal nerve system of the Bobolink (*Dolichonyx oryzivorus*). *Neurosci. Lett.* **80**, 229–234 (1987).
21. Hellmann, B. & Güntürkün, O. Structural organization of parallel information processing within the tectofugal visual system of the pigeon. *J. Comp. Neurol.* **429**, 94–112 (2001).
22. Diekamp, B., Hellmann, B., Troje, N. F., Wang, S. R. & Güntürkün, O. Electrophysiological and anatomical evidence for a direct projection from the nucleus of the basal optic root to the nucleus rotundus in pigeons. *Neurosci. Lett.* **305**, 103–106 (2001).
23. Mai, J. K. & Semm, P. Pattern of brain glucose utilization following magnetic stimulation. *J. Hirnforsch.* **31**, 331–336 (1990).
24. Güntürkün, O. & Hahmann, U. Functional subdivisions of the ascending visual pathways in the pigeon. *Behav. Brain Res.* **98**, 193–201 (1999).
25. Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1981).

Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft. We thank K. Dutine, F. Eich, F. Schmidt and A. Wittekindt for their help with conducting the experiments.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.W. (e-mail: wiltchko@zoology.uni-frankfurt.de).

Moving visual stimuli rapidly induce direction sensitivity of developing tectal neurons

Florian Engert*†‡, Huizhong W. Tao*†, Li I. Zhang†§ & Mu-ming Poo*

* Division of Neurobiology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

§ Keck Center of Integrative Neuroscience, University of California, San Francisco, California 94143, USA

† These authors contributed equally to this work

During development of the visual system, the pattern of visual inputs may have an instructive role in refining developing neural circuits^{1–4}. How visual inputs of specific spatiotemporal patterns shape the circuit development remains largely unknown. We report here that, in the developing *Xenopus* retinotectal system, the receptive field of tectal neurons can be 'trained' to become direction-sensitive within minutes after repetitive exposure of the retina to moving bars in a particular direction. The induction of direction-sensitivity depends on the speed of the moving bar, can not be induced by random visual stimuli, and is accompanied by an asymmetric modification of the tectal neuron's receptive field. Furthermore, such training-induced changes require spiking of the tectal neuron and activation of a NMDA (*N*-methyl-D-aspartate) subtype of glutamate receptors during training, and are attributable to an activity-induced enhancement of glutamate-mediated inputs. Thus, developing neural circuits can be modified rapidly and specifically by visual inputs of defined spatiotemporal patterns, in a manner consistent with predictions based on spike-time-dependent synaptic modification.

Spontaneous and experience-evoked activities in the developing brain can influence the refinement of developing nerve connections into mature neural circuits. In the visual system, rearing kittens with

an artificial squint leads to failure in the development of binocular response properties of striate cortex neurons⁵. Blockade of spontaneous waves of retinal activity also disrupts eye-specific segregation of retinal inputs to the lateral geniculate nucleus^{6,7}. Synchronizing retinal inputs by strobe light or electrical stimulation affects formation of normal receptive field properties in various systems^{8–10}. Furthermore, an instructive role of visual inputs is indicated by the appearance of visual modules in the auditory cortex of the ferret after rewiring of retinal inputs^{11,12}. In the present study, we found a rapid and specific modification of receptive field properties of tectal neurons by the visual input of a defined spatiotemporal pattern, in a manner consistent with hebbian synaptic modification as a mechanism for activity-dependent changes in visual circuits^{13,14}.

The effect of visual experience on the receptive field properties of tectal neurons was examined in developing *Xenopus* tadpoles. Patterned visual inputs were used to stimulate the retina, and tectal cell responses were monitored with *in vivo* perforated whole-cell recording methods (Fig. 1a). First, we mapped the receptive field of the tectal neuron by random and sequential flashing of a white square at various locations on the retina (see Methods). The integrated charge of stimulus-evoked compound synaptic currents (CSCs) was measured within a defined window for the more prominent 'off' responses (Fig. 1b). The measured value at each location was represented in grey scale as one element of an 8 × 7 grid that covered the total area of the projected visual image on the retina (Fig. 1c). This analysis based on CSCs reveals a large receptive field (50–80% of the retina, *n* = 12) of tectal neurons at these early stages (42–45), consistent with a diffuse retinotectal connectivity during early development^{15,16}.

To examine the effect of patterned visual inputs on the receptive field property of tectal neurons, we stimulated the retina with white moving bars (20-μm wide, speed 0.3 μm ms⁻¹) in four orthogonal directions (right, down, left and up, Fig. 2a) and recorded the responses of tectal cells evoked by moving visual stimuli. In all tectal cell responses (voltage-clamped at -70 mV, *n* = 20), we found no

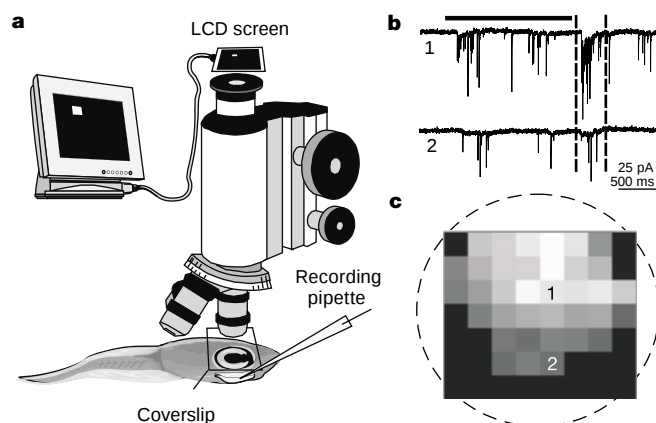


Figure 1 Mapping the receptive field of developing *Xenopus* tectal neurons. **a**, Diagram of the experimental set-up, depicting the recording of a tectal neuron in the exposed tectum. **b**, Traces shown are samples of compound synaptic currents (CSCs) evoked by the flashing of a white square (for 1.5 s, bar) within the receptive field at two locations (marked 1 and 2). Recording was made in voltage clamp at the reversal potential of Cl⁻ current (-45 mV) to reveal glutamate-mediated inputs. **c**, The receptive field was assayed by measuring the integrated charge of the 'off' response of CSCs within a defined window (dotted lines in **b**). The value of CSC charge is represented by linear grey scale (black = average basal activity without light stimulus) as a corresponding element in the 8 × 7 grid covering the projection area. The dashed circle marks the approximate position of the retina.

‡ Present address: Department of Molecular and Cell Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

apparent directional preference to moving bars in any of the four directions, in terms of the total charge of CSCs evoked by each moving stimulus (see Fig. 2b, green bars), although these CSCs exhibited a distinct profile for each direction (Fig. 2a). As these developing tectal neurons had no apparent directional preference, we inquired whether repeated exposure of directional stimuli can 'train' these neurons to acquire direction-sensitivity in their responses. After the initial assay of tectal cell responses to moving-bar stimuli in each of the four directions (voltage-clamped at -70 mV), the recording was switched to current clamp and the retina was exposed to 60 sweeps of the moving bar (speed $0.3 \mu\text{m ms}^{-1}$, frequency 0.2 Hz) in one specific, but randomly chosen, direction (rightward in Fig. 2). In many neurons, each sweep elicited reliably a train of spikes (Fig. 2b, inset). When the tectal cell responses were assayed again after the training, we found that the response to the training stimulus (bar to the right in Fig. 2a) was enhanced to a level significantly larger than those for other directions. In Fig. 2a, the effect is further illustrated by superimposing the current traces before and after training for each direction. The enhancement of CSCs also led to an increased

spike rate of the tectal neuron in response to the training stimulus (Fig. 2c). In most experiments, we monitored CSCs (in voltage clamp) instead of spiking activities to avoid spiking-dependent changes caused by testing. Figure 2d depicts the time course of training-induced changes in seven experiments, for which the training stimulus reliably induced spiking of the tectal cell. The enhancement of tectal cell response to the test stimulus of the trained direction was persistent for up to 50 min.

In many regions of the developing brain^{14,17–20}, the timing of spiking in pre- and postsynaptic neurons is critical for the induction of long-term potentiation (LTP) and long-term depression (LTD). Modelling studies indicate that such spike-time-dependent plasticity may provide a mechanism for the development of direction-selectivity of visual circuits for detecting moving stimuli^{21–24}. Moving bars probably evoke consecutive spiking of neighbouring retinal ganglion cells and of the tectal cells they innervate, with specific temporal relationships. Such a spatiotemporal pattern of spiking may be essential for triggering synaptic changes underlying the development of direction-selectivity. We further tested this idea by using bar stimuli that moved at slower speeds. As shown in Fig. 3a,

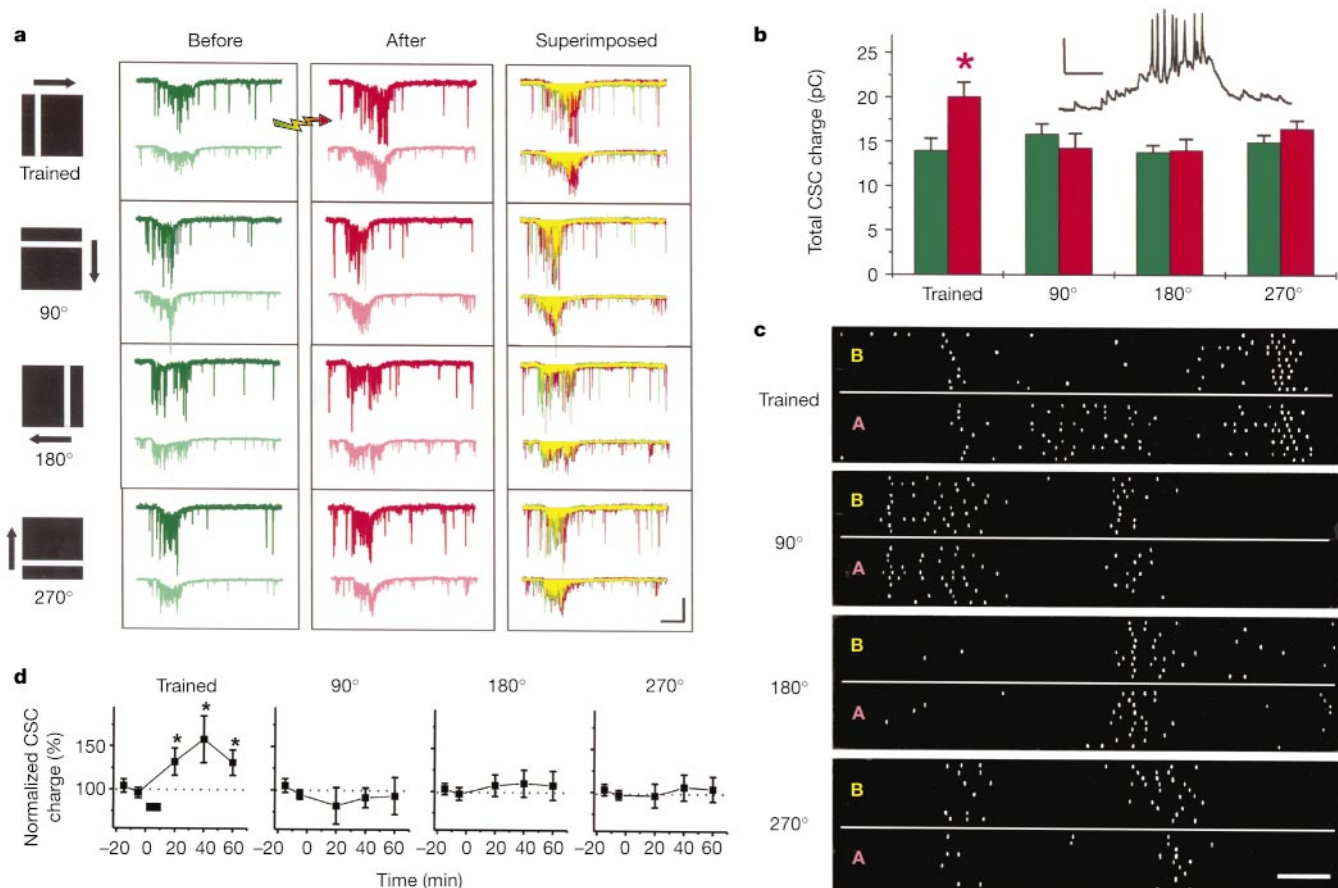


Figure 2 Selective enhancement of tectal cell responses by training with a moving bar. **a**, Traces of CSCs evoked by test moving bars of four orthogonal directions before (green) and after (red) repetitive exposure of the retina to rightward moving bars (coloured arrow). Bright traces indicate CSCs evoked by a single test sweep; dim traces indicate averages of three CSCs. Superimposed traces reveal the enhanced tectal response for the trained direction (see red area). Scale bar: 50 pA (vertical), 500 ms (horizontal). **b**, Average total CSC charge ($\pm$ s.e.m., $n = 6$) evoked by the moving stimulus before (green) and after (red) training, for the experiment shown in **a**. The average CSC charge for the trained direction was significantly enhanced, as compared with either its value before training or after-training values for all three other directions (asterisk, $P < 0.02$, two

tailed t -test). Inset, sample trace of tectal cell response evoked by the training stimulus (in current clamp, $V_m = -50$ mV). Scale bar: 20 mV (vertical), 300 ms (horizontal). **c**, Training-induced changes in the stimulus-evoked spiking activity of the tectal neuron. Raster plots show spike trains evoked by ten consecutive test moving stimuli before (B) and after (A) the training (as in **a**), with each white dot depicting a spike. Scale bar, 100 ms. **d**, Persistency of training-induced, direction-sensitive tectal responses ($n = 7$). Training was performed at 0–10 min (black bar). Data from each experiment were normalized to the average pre-training control and pooled into five bins. Asterisk, significantly different from the corresponding data for other directions ($P < 0.05$, paired t -test).

no significant direction-specific change was induced by training with the slow-moving bar ($0.1 \mu\text{m ms}^{-1}$), whereas bars moving with a medium speed ($0.2 \mu\text{m ms}^{-1}$) induced a slight increase in the tectal response to the training stimulus. Comparing the patterns of tectal spikes, we found that the fast bar generated spikes with significantly shorter inter-spike intervals (Fig. 3d), although the total number of spikes evoked by each sweep was the same (5.4 ± 0.8 and 5.6 ± 0.7 for the fast and slow sweeps, respectively). Thus the timing of sequential excitation of retinal cells seems to be critical for inducing direction-specific changes in the tectal neurons. In addition, we found that after training with the fast-moving bar, the tectal response became enhanced only for the test bar that moved at the fast but not the slow speed (Fig. 3b). Thus the circuit modification depends on the speed of the training stimulus, and the modified circuit is specific for the detection of the training stimulus.

The importance of the spatiotemporal pattern of retinal excitation in the observed training effects was further examined by exposing the retina to randomized training stimuli. We used non-overlapping white squares that flashed in random sequence and covered the same area of the visual field for the same amount of time as the fast-moving bar stimulus (Fig. 3c). Each set of random flashes evoked a similar total number of spikes (5.3 ± 0.7) as that by a sweep of the fast-moving bar, but the distribution of inter-spike intervals was significantly different, with spikes dispersed over longer intervals (Fig. 3d). After training with 120 sets of random flash stimuli, we found no enhancement of tectal responses in any direction (Fig. 3c). Thus the pattern of spiking, presumably in both pre- and postsynaptic neurons, is critical in the observed induction of direction-sensitive responses. Finally, we tested the effect of training with bars of four orthogonal directions, with a random sequence of occurrence but the same number (60) of sweeps for each direction. We found that the tectal responses after the training were enhanced in all four directions (Fig. 3c), although the level of enhancement was lower than that induced by the unidirectional training. Thus the tectal cell is not pre-determined to be direction-

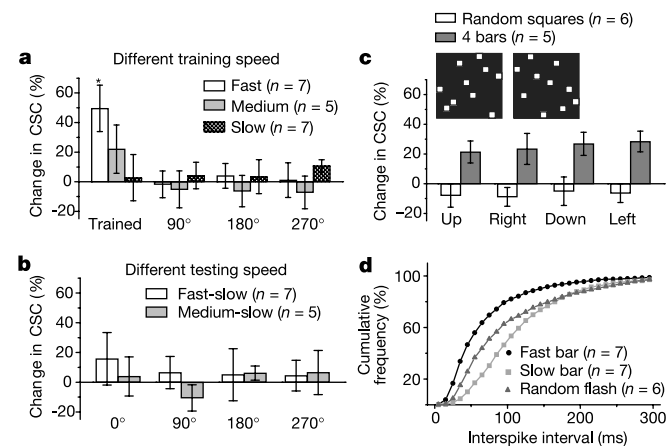


Figure 3 Effect of training with stimuli of different spatiotemporal patterns. **a**, Percentage changes in the total CSC charge for responses to moving bars in four test directions after training with three different speeds. Data (mean $\pm$ s.e.m.) within 30 min after training were averaged and normalized to the pre-training control. The same speed of the moving bar was used for training and testing. Asterisk, significantly different from other test directions ($P < 0.01$, t -test). **b**, Normalized responses to slow test bars after training with the fast (fast-slow) or the medium-speed (medium-slow) moving bar, respectively. 0°, test direction the same as the training direction. **c**, Results of training with random flashing squares (white, see insets for two examples of the image) and with four orthogonal moving bars in random sequence (grey). **d**, Cumulative distribution of inter-spike intervals for tectal cell spikes during training in response to a sweep of the fast- or slow-moving bar, or an equivalent set of random flashing squares.

selective in a particular direction. The appearance of direction-selectivity *in vivo* presumably requires an imbalance of moving visual stimuli and other mechanisms for consolidating the acquired response properties of the tectum.

To explore cellular changes induced by the training stimuli, we examined further the receptive fields of tectal cells before and after training. To facilitate rapid mapping, we used a set of ten vertical and ten horizontal flashing bars in random sequence to determine the receptive field profiles of the tectal neuron in horizontal and vertical directions, respectively (Fig. 4a). After training with fast bars in one direction, an asymmetric change in the profile of the tectal receptive field was revealed by the enhanced tectal cell responses only for upstream test bar locations (Fig. 4a). As sum-

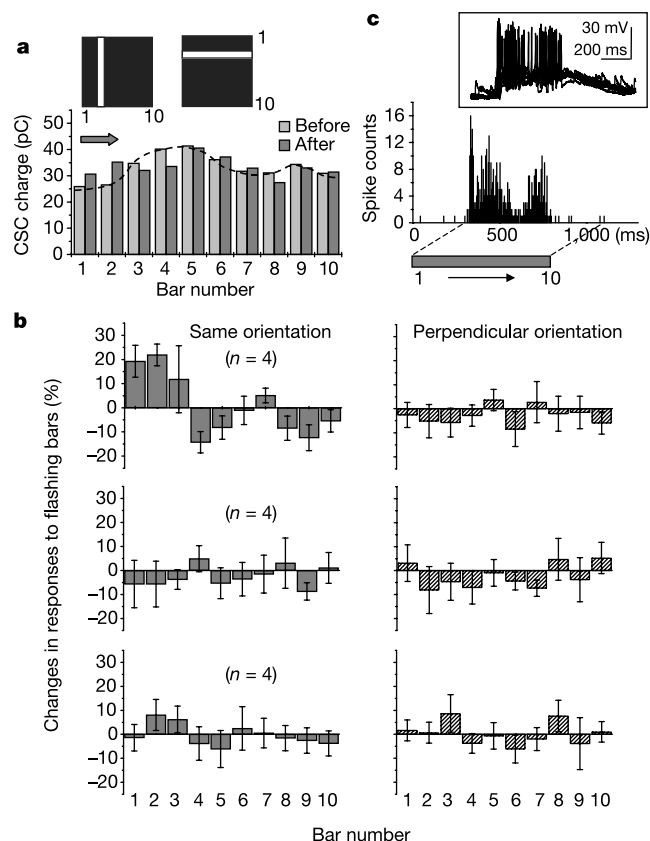


Figure 4 Asymmetric modification of the tectal receptive field by training with moving stimuli. **a**, Average CSC charge of tectal cell responses (eight repeats) to a set of ten vertical flashing bars (labelled 1–10 along the training direction, each flashed for 1.5 s, left inset) before and after training with fast (rightward, arrow) moving bars. Dashed curve, receptive field profile before training. The receptive field profile was also determined for the perpendicular dimension by measuring the tectal responses to horizontal bars (numbered 1–10 along the direction that was 90° clockwise to the training direction, right inset box; data not shown). Recordings were made at the reversal potential of Cl^- currents. **b**, Top panel, summary of four experiments similar to **a**. Histograms depict the mean percentage changes after training in tectal responses to flashing bars of the same (grey) and perpendicular (hatched) orientation with respect to the training direction. The middle panel is the same as the top panel except that the tectal cell was hyperpolarized constantly (at -80 mV in current clamp) and prevented from spiking during training. The bottom panel is the same as the top panel except that the training stimulus was the slow-moving bar. **c**, Post-stimulus spike-timing histogram for tectal responses during the training for the experiment shown in **a**. The onset of the moving stimulus on the screen was set as time 0, and it took about 700 ms for the bar to sweep across the visual field (bar below). The left dashed line marks the onset of the stimulus-evoked synaptic potential with respect to the onset of the stimulus. Inset, superimposed tectal potentials (six samples) during training.

marized in Fig. 4b (top panel), asymmetric receptive field modification was induced by fast-moving bars along the direction of the movement, whereas no apparent asymmetry was found for receptive field in the perpendicular dimension. This is analogous to the asymmetric change in hippocampal place field induced by unidirectional movement of a rat in a closed track²⁵, which is consistent with the prediction by a computational model based on asymmetric spike-time-dependent synaptic modification²⁶. Spiking of tectal neurons evoked by the fast-moving bar appears with a delay of a few hundred ms after the onset of the stimulus sweep on the projection screen (Fig. 4c), with most spikes clustered during the early phase of the sweep, where modification of the receptive field was most prominent. In addition, consistent with the involvement of spike-time-dependent plasticity, no change in the receptive field profile was observed when the tectal cell was hyperpolarized to be prevented from spiking during training (Fig. 4b, middle panel), or

when the slow-moving bar was used for training (Fig. 4b, bottom panel).

Direction-sensitive tectal responses induced by training may result from changes in the tectum or in the retina. We found that when unidirectional fast-bar training was performed under voltage clamp (-80 mV) of the tectal neuron, there was no significant change in the responses of tectal neurons to stimuli in any of the four directions (Fig. 5a). In three experiments, two training sessions were given sequentially, with the tectal cell in voltage clamp (-80 mV) and then in current clamp. Post-training tests showed no change in the tectal responses after training in voltage clamp ($-10\% \pm 10$, s.e.m.), but a significant enhancement after training in current clamp ($48\% \pm 11\%$). Thus the induction of direction-sensitive responses required depolarization of the tectal cell. Furthermore, training stimuli that consistently failed to elicit spike trains in the tectal cell did not induce responses with

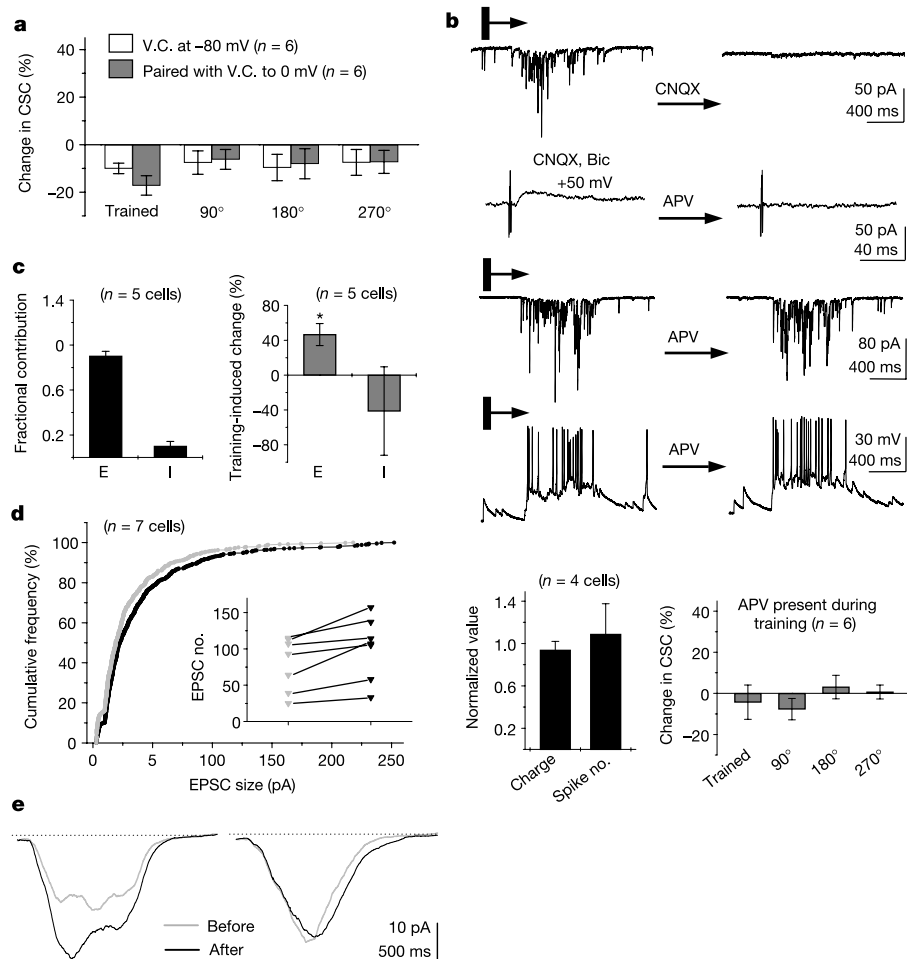


Figure 5 Synaptic mechanisms. **a**, Experiments similar to Fig. 2a except that during training the tectal cell was voltage-clamped (v.c.) at -80 mV (white) or depolarized from -70 to 0 mV for 700 ms during each training sweep (grey). **b**, The top upper panel shows sample traces of CSCs evoked by a moving bar before and after local perfusion of the tectum with CNQX. The top lower panel shows NMDA receptor-mediated synaptic currents (voltage-clamped at $+50$ mV) induced by electrical stimulation of the retina (with CNQX and bicuculline in the bath) before and after local perfusion with D-APV. The middle panel shows sample traces of moving-bar-induced CSCs and spikes before and after local perfusion with D-APV. The bottom left panel shows the average total CSC charge and spike number evoked by the moving-bar stimulus after D-APV perfusion, normalized for each cell to the pre-perfusion value. The bottom right panel shows the average percentage changes ($\pm$ s.e.m.) in the CSC charge of responses to fast-moving bars after

training, with D-APV present during training. **c**, Fractional contribution of Na^+ (excitatory (E)) and Cl^- currents (inhibitory (I)) in the CSCs evoked by the moving-bar stimulus (left). Training-induced changes in the 'E' and 'I' components of CSCs (asterisk, $P < 0.01$, t -test) (right). **d**, Cumulative distribution of the amplitude of individual EPSCs associated with the CSCs evoked by the moving bar of the trained direction before (light) and after (dark) training. The two curves are significantly different ($P < 0.01$, Kolmogorov-Smirnov test). Inset, the total number of EPSCs associated with the same CSCs before (grey) and after (black) training. Lines connect data points from the same experiment. Two groups are significantly different ($P < 0.01$, paired t -test). **e**, Temporal profile of CSCs evoked by moving stimulus of the trained (left) and the opposite direction (right) before and after training. Trace represents the average of CSCs from seven experiments, as in Fig. 2d.

significant directional sensitivity ($n = 9$, data not shown). The requirement of postsynaptic depolarization is reminiscent of hebbian LTP observed at many central synapses^{27,28}, including LTP of these retinotectal synapses induced by direct electrical stimulation of the retina¹⁸ or by dimming light stimuli²⁹. However, step depolarization of the tectal cell (voltage-clamped from -70 to 0 mV, 700 ms), when paired with each training stimulus, was not sufficient to enhance the response to the training stimulus (Fig. 5a). Thus postsynaptic spike-associated depolarization and perhaps other factor(s) are essential for the observed training effect.

In many regions of the brain^{27,28}, including the developing tectum¹⁸, activity-induced LTP requires activation of NMDA receptors. To test further whether the mechanism underlying changes to the receptive field described here involves NMDA receptor-dependent synaptic plasticity in the tectum, we perfused the tectum locally with the NMDA receptor antagonist D-2-amino-5-phosphopentanoic acid (APV). The effectiveness of the local perfusion of drugs was shown by the finding that perfusion with CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) abolished fast transient currents associated with moving-bar-induced CSCs, and that perfusion with APV blocked NMDA receptor-mediated synaptic currents induced by electrical stimulation of retinal ganglion cells (Fig. 5b, top panel). Perfusion of APV did not affect significantly the tectal response to each sweep of the moving bar, in terms of either total CSC charge or spiking activity (Fig. 5b, middle and bottom left panels), but it abolished completely direction-specific change in the tectal responses induced by training (Fig. 5b, bottom right panel). Thus, activation of NMDA receptors is required for the induction of direction-sensitive responses, consistent with the blocking effect of APV on LTP of retinotectal synapses¹⁸.

Further analysis was carried out to determine the relative contributions of excitatory and inhibitory inputs to the CSCs evoked by each moving-bar stimulus. On the basis of measurements of the average CSC charge at two different clamping voltages and the reversal potentials for Na^+ and Cl^- currents (see Methods), we found that excitatory glutamate-mediated inputs contributed to about 90% of the total CSC charge at -70 mV (Fig. 5c, left). After training with fast bars, there was selective increase in these excitatory inputs in terms of the total charge of excitatory components of the CSCs (Fig. 5c, right), as well as a significant increase in both the size and the average frequency of excitatory postsynaptic currents (EPSCs) associated with CSCs evoked by test bars of the training direction (Fig. 5d). Taken together, these results suggest that the training had induced a potentiation of glutamate-mediated inputs on these tectal cells.

The observed induction of direction-sensitivity of tectal neurons could result from a strengthening of retinotectal connections made by direction-selective retinal ganglion cells at this early stage of development. Alternatively, it may arise from a training-induced circuit modification within the tectum. In principle, direction-selectivity of a visual neuron can develop through an activity-dependent asymmetric circuit modification that results in excitation of the neuron only when the stimulus of the preferred direction is presented^{21–24}. We note that, after training with the moving bar, the profiles of CSCs were asymmetrically altered for responses elicited by stimuli in the trained and the opposite directions (Fig. 5e), suggesting distinct circuit modifications for detection of moving signals in preferred and non-preferred directions. The observed requirement of tectal-cell spiking is consistent with a circuit modification that involves spike-time-dependent modification of retinotectal and/or intra-tectal connections, although training-induced changes in the retina may also occur. Although the precise loci of circuit changes remain to be determined, the rapidity and persistence of receptive field modification shown here illustrate the susceptibility of developing neural circuits to the instructive influence by sensory inputs of specific spatiotemporal patterns. □

Methods

Tadpole preparation and electrophysiology

Xenopus laevis tadpoles of Nieuwkoop and Faber stage 42–45 were anaesthetized with saline containing 0.02% MS222 (Sigma) and secured by insect pins to a sylgard-coated dish, and incubated in HEPES-buffered saline containing (in mM): 115 NaCl, 2 KCl, 10 HEPES, 3 CaCl_2 , 10 glucose, 1.5 MgCl_2 , and 0.005 glycine (pH 7.4). For recording, the skin was removed and the brain was split open along the midline to expose the inner surface of the tectum. A low dose of α -bungarotoxin (2 mg ml^{-1}) was applied to the bath to prevent muscle contraction. As shown previously¹⁸, this toxin treatment did not significantly affect the retinotectal responses. The method of perforated-patch, whole-cell recording has been described previously³⁰. The micropipettes were made from borosilicate glass capillaries (Kimax), had a resistance in the range of 3 – 4 M Ω , and were tip-filled with internal solution and then back-filled with internal solution containing 200 mg ml^{-1} amphotericin B. The internal solution contained (in mM): 110 K-gluconate, 10 KCl, 5 NaCl, 1.5 MgCl_2 , 20 HEPES, 0.5 EGTA (pH 7.3). Experiments were performed at room temperature (22°C) and the bath was constantly perfused with fresh recording medium at a slow rate (1 ml min^{-1}). Recording was made with a patch-clamp amplifier (Axopatch 200A; Axon Instruments). The whole-cell capacitance was fully compensated and the series resistance (10 – 20 M Ω) was compensated at 75 – 80% (lag 60 μs). Signals were filtered at 5 kHz and sampled at 10 kHz using Axoscope software (Axon Instruments). Local perfusion of the tectum was carried out with a glass pipette (opening 30 – 40 μm) placed near the tectum. All drugs were from Sigma. Concentrations used were: 10 μM CNQX, 10 μM bicuculline, 50 μM D-APV.

For mapping of the receptive field, recordings were made under the reversal potential for Cl^- current ($E_i = -45$ to -60 mV), which was determined by the disappearance and the reversal of spontaneous GABA (γ -aminobutyric acid)-mediated synaptic currents as the holding potential was changed towards more depolarized values. To determine the contribution of excitatory and inhibitory components of stimulus-evoked CSCs, recordings were repeated at two holding potentials (-90 and -70 mV). The Na^+ and Cl^- conductances were determined by the formula $I(t) = (G_e + g_e(t))(V - E_e) + (G_i + g_i(t))(V - E_i)$, where $I(t)$ is the amplitude of CSCs at time t ; $g_e(t)$ and $g_i(t)$ are Na^+ and Cl^- conductances; V is the clamping voltage; E_e and E_i are reversal potentials for Na^+ and Cl^- currents, respectively; and G_e and G_i are leakage conductances (which are negligible). Measurements of $I(t)$ at two voltages (averages of 20 repeats) yielded $g_e(t)$ and $g_i(t)$, and the relative contribution of Na^+ and Cl^- currents to CSCs were calculated. For estimates of changes in EPSCs, fast transients of CSCs were identified by custom-made LabVIEW (National Instruments) software to determine the peak amplitude of individual EPSCs and their frequency. For analysis of spiking patterns, the peak of spikes observed in the tectal responses was identified and the inter-spike intervals were calculated.

Visual stimulation

The tectal cell was patched under visual control and the retina was flattened and stabilized with a glass coverslip after removal of the lens. A small LCD screen (from a virtual reality goggles, Sony, PLM-A35) was mounted on the camera port of the microscope, allowing projection of computer-generated images onto the retina (Fig. 1a). The diameter of the retina was in the range of 250 – 300 μm and that of the tectal receptive field in the range of 100 – 200 μm . Visual stimulation and analysis software was custom-made. For receptive field mapping, the entire field for image projection (200×200 μm) was divided into a 8×7 grid. White squares (corresponding to each element of the grid) were flashed for 1.5 s in a random sequence, with 5 -s intervals. For moving-bar stimuli, white bars (20 μm in width at the retina) swept across the screen at a speed of 0.3 , 0.2 or 0.1 $\mu\text{m ms}^{-1}$ (fast, medium and slow speed, respectively). For testing tectal responses, bars moving in four orthogonal directions (in the sequence of up, right, down and left) were presented with 10 -s intervals, with 3 – 5 repeats in each testing session. Two or three testing sessions (with 5 -min interval) were performed during the control period. In some experiments, responses to two different speeds were tested. The training session consisted of 60 sweeps of a randomly chosen direction (0.2 Hz), or 240 sweeps of four orthogonal directions with a randomized order (0.5 Hz). For random flashing stimulus, white squares of the same width as the moving bar were flashed at random at non-overlapping locations (see example images in Fig. 3c, inset), exposing the retina to the same total amount of light at any given moment, and covering the same total area as the fast-moving bar stimulus. A total of 120 sets of stimuli (0.2 Hz) were applied for training. For rapid mapping of training-induced changes in the receptive field, the image projection field was divided into a set of ten horizontal and ten vertical bars. Each bar was flashed for 1.5 s (at 5 -s interval) and the sequence of flashing bars was randomized. After preliminary mapping of the receptive field, the size of the image projection field was sometimes slightly re-adjusted to cover the entire receptive field. The receptive field was then mapped in two sessions (four repeats in each session) before and after training.

Received 6 May; accepted 12 July 2002; doi:10.1038/nature00988.

- Wiesel, T. N. Postnatal development of the visual cortex and the influence of environment. *Nature* **299**, 583–591 (1982).
- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Penn, A. A. & Shatz, C. J. Brain waves and brain wiring: the role of endogenous and sensory-driven neural activity in development. *Pediatr. Res.* **45**, 447–458 (1999).
- Zhang, L. I. & Poo, M. M. Electrical activity and development of neural circuits. *Nature Neurosci.* **4** Suppl., 1207–1214 (2001).
- Hubel, D. H. & Wiesel, T. N. Binocular interaction in striate cortex of kittens reared with artificial stimuli. *J. Neurophysiol.* **28**, 1041–1059 (1965).
- Meister, M., Wong, R. O., Baylor, D. A. & Shatz, C. J. Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science* **252**, 939–943 (1991).

7. Penn, A. A., Riquelme, P. A., Feller, M. B. & Shatz, C. J. Competition in retinogeniculate patterning driven by spontaneous activity. *Science* **279**, 2108–2112 (1998).
8. Cynader, M., Berman, N. & Hein, A. Cats reared in stroboscopic illumination: effects on receptive fields in visual cortex. *Proc. Natl Acad. Sci. USA* **70**, 1353–1354 (1973).
9. Schmidt, J. T. & Eisele, L. E. Stroboscopic illumination and dark rearing block the sharpening of the regenerated retinotectal map in goldfish. *Neuroscience* **14**, 535–546 (1985).
10. Weliky, M. & Katz, L. C. Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature* **386**, 680–685 (1997).
11. Sharma, J., Angelucci, A. & Sur, M. Induction of visual orientation modules in auditory cortex. *Nature* **404**, 841–847 (2000).
12. von Melchner, L., Pallas, S. L. & Sur, M. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature* **404**, 871–876 (2000).
13. Fregnac, Y., Shulz, D., Thorpe, S. & Bienenstock, E. A cellular analogue of visual cortical plasticity. *Nature* **333**, 367–370 (1988).
14. Schuett, S., Bonhoeffer, T. & Hubener, M. Pairing-induced changes of orientation maps in cat visual cortex. *Neuron* **32**, 325–337 (2001).
15. Gaze, R. M., Keating, M. J. & Chung, S. H. The evolution of the retinotectal map during development in *Xenopus*. *Proc. R. Soc. London B* **185**, 301–330 (1974).
16. Holt, C. E. & Harris, W. A. Order in the initial retinotectal map in *Xenopus*: a new technique for labelling growing nerve fibres. *Nature* **301**, 150–152 (1983).
17. Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
18. Zhang, L., Tao, H. W., Holt, C. E., Harris, W. A. & Poo, M. M. A critical window for cooperation and competition among developing retinotectal synapses. *Nature* **395**, 37–44 (1998).
19. Feldman, D. E. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron* **27**, 45–56 (2000).
20. Boettiger, C. A. & Doupe, A. J. Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* **31**, 809–818 (2001).
21. Abbott, L. F. & Blum, K. I. Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex* **6**, 406–416 (1996).
22. Zanker, J. M. & Zeil, J. *Motion Vision—Computational, Neural, and Ecological Constraints* (Springer, New York, 2000).
23. Rao, R. P. & Sejnowski, T. J. Predictive learning of temporal sequences in recurrent neocortical circuits. *Novartis Found. Symp.* **239**, 208–229 (2001).
24. Roberts, P. D. Computational consequences of temporally asymmetric learning rules. I. Differential hebbian learning. *J. Comput. Neurosci.* **7**, 235–246 (1999).
25. Mehta, M. R., Barnes, C. A. & McNaughton, B. L. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921 (1997).
26. Mehta, M. R., Quirk, M. C. & Wilson, M. A. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715 (2000).
27. Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
28. Malenka, R. C. & Nicoll, R. A. Long-term potentiation—a decade of progress? *Science* **285**, 1870–1874 (1999).
29. Zhang, L., Tao, H. & Poo, M. Visual input induces long-term potentiation of developing retinotectal synapses. *Nature Neurosci.* **3**, 708–715 (2000).
30. Rae, J., Cooper, K., Gates, P. & Watsky, M. Low access resistance perforated patch recordings using amphotericin B. *J. Neurosci. Methods* **37**, 15–26 (1991).

Acknowledgements

This work was supported by grants from NSF and NIH. F.E. was supported in part by a long-term fellowship from the Human Frontier Science Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.-m.P. (e-mail: mpoo@uclink.berkeley.edu).

Dendrite growth increased by visual activity requires NMDA receptor and Rho GTPases

Wun Chey Sin, Kurt Haas, Edward S. Ruthazer & Hollis T. Cline

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

Previous studies suggest that neuronal activity may guide the development of synaptic connections in the central nervous system through mechanisms involving glutamate receptors and GTPase-dependent modulation of the actin cytoskeleton^{1–7}. Here we demonstrate by *in vivo* time-lapse imaging of optic tectal cells

in *Xenopus laevis* tadpoles that enhanced visual activity driven by a light stimulus promotes dendritic arbor growth. The stimulus-induced dendritic arbor growth requires glutamate-receptor-mediated synaptic transmission, decreased RhoA activity and increased Rac and Cdc42 activity. The results delineate a role for Rho GTPases in the structural plasticity driven by visual stimulation *in vivo*.

We used *in vivo* time-lapse imaging of single optic tectal neurons in *Xenopus* tadpoles to test the function of visual activity in neuronal development. We compared the dendritic arbor growth rates of individual tectal neurons during a 4-h period with a visual stimulus to a preceding 4-h period in the absence of light. This imaging protocol allows the comparison of dendritic arbor structures of the same neurons over time and therefore provides a sensitive measure of structural plasticity. Visual stimulation significantly enhanced dendritic arbor elaboration compared with growth rates in the preceding 4-h period in the dark (Fig. 1a, b; see also Supplementary Table 1). Neurons from animals exposed to visual stimulus throughout the 8-h protocol maintained a constant rate of dendritic arbor elaboration (Fig. 1b; see also Supplementary Fig. 1 and Table 1). This indicates that growth rates do not change with longer periods of stimulation. Another set of animals was exposed to visual stimulus within the first 4-h period and then returned to the dark

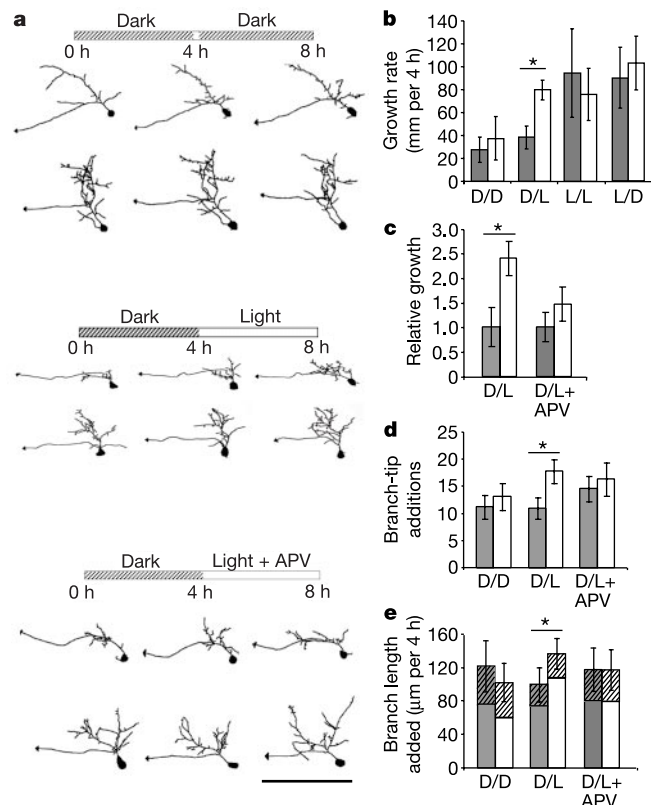


Figure 1 Visual stimulation *in vivo* promotes dendritic arbor growth by a glutamate-receptor-dependent mechanism. **a**, Drawings of neurons imaged three times at 4-h intervals. Two examples are shown for each treatment. Animals were placed in a dark chamber for 4 h (dark) or a chamber with a light stimulus for 4 h (light) in the presence or absence of APV as depicted by the bar over the neurons. Arrowheads identify efferent axons in this and subsequent figures. Scale bar, 100 μm. **b**, Quantification of dendritic arbor growth rates during 4 h in the dark (D) or with visual stimulation (L). **c**, Dendritic arbor growth rates normalized to the growth rate in the 0–4 h period in the dark (D). **d**, Quantification of branch-tip additions. **e**, Contribution of new branches or branch extension (hatched, top region) to increased branch length. Asterisk, $P < 0.05$.

for the second 4-h period. Notably, the enhanced growth rates of these neurons in response to light stimulation were maintained during the 4-h dark period (Fig. 1b; see also Supplementary Fig. 1). This suggests that exposure to visual stimulation triggers mechanisms that enhance dendritic arbor growth over a period following termination of the stimulus.

Although the experimental protocol was designed to detect

structural changes within individual neurons, we also detect a significant increase in the growth rate of neurons from animals exposed to visual stimulation during the first 4 h of the experiment ($68.3 \pm 12.1 \mu\text{m}$, $n = 16$) compared with the growth rate of neurons from animals placed in the dark for the initial 4 h ($35.4 \pm 6.0 \mu\text{m}$, $n = 84$, $P < 0.002$). This indicates that the increased growth rate is due to the visual stimulus rather than

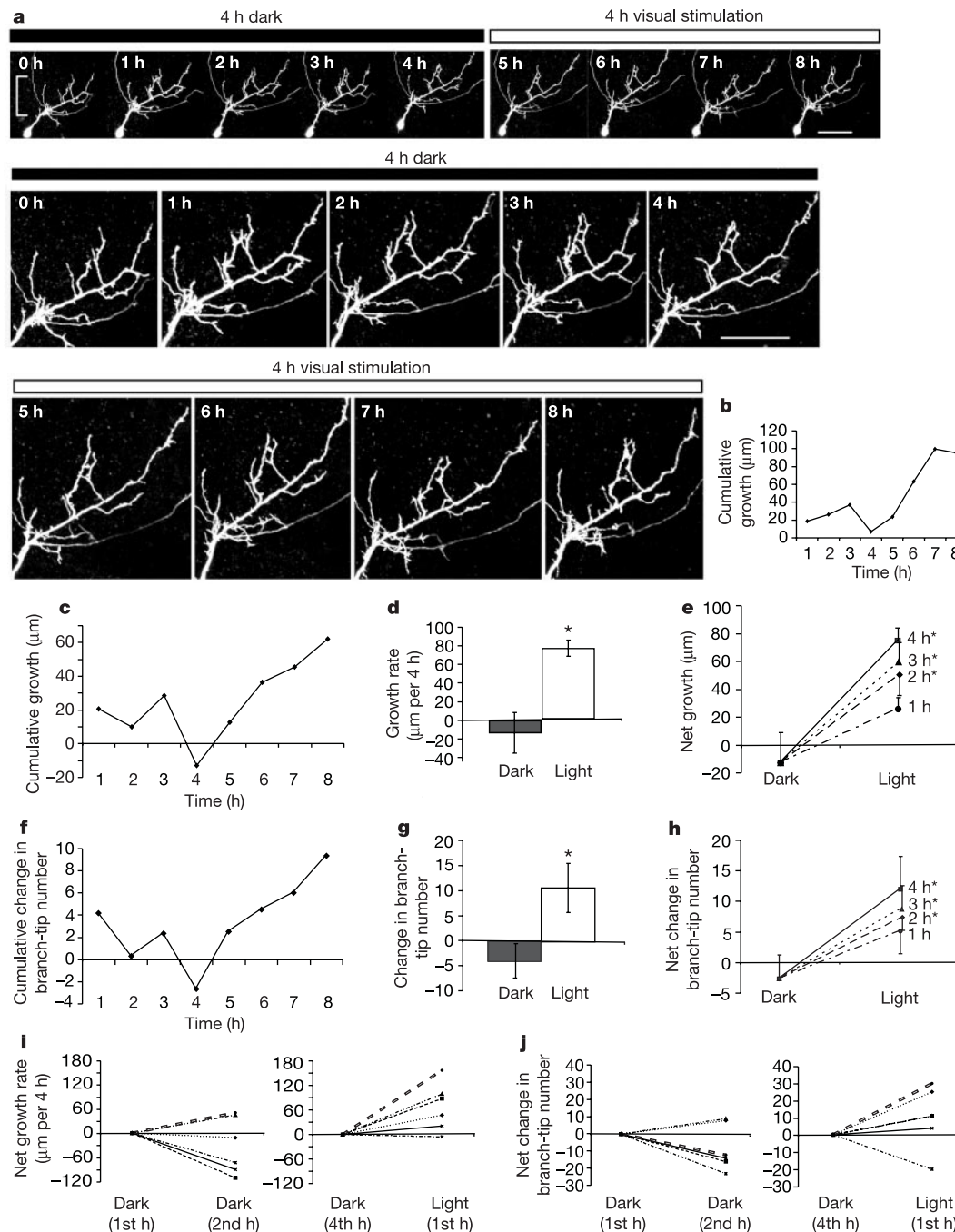


Figure 2 Visual system activity regulates branch dynamics *in vivo*. **a**, Two-photon images of a neuron collected once an hour over 8 h, during a visual stimulus regime depicted by the bars over the images. The bottom two panels show greater magnification of the bracketed region in the top panel. Scale bars, $50 \mu\text{m}$. **b**, Cumulative change in branch length for the neuron in **a**. **c**, **f**, Average cumulative change in branch length (**c**) and branch-tip number (**f**) ($n = 6$); **d**, **g**, Average change in branch length (**d**) and branch-tip number (**g**) over the 4-h period in the dark (dark) or with visual stimulus (light).

e, **h**, Cumulative change in branch length (**e**) and branch-tip number (**h**) with each hour of visual stimulus (light) compared with 4 h without visual stimulus (dark). **i**, **j**, Changes in branch length (**i**) and branch-tip number (**j**) for individual neurons from the first to second hours in the dark (left graphs) compared with changes during the last hour in the dark and first hour with visual stimulation (right graphs). Asterisk, $P < 0.05$.

a reaction to the 4 h in the dark.

The light-induced increase in dendritic arbor growth rate was blocked by exposure to APV (3-amino-phosphonovaleric acid; Fig. 1a, c) and CNQX (6-cyano-nitroquinoxaline-2,3-dione; Supplementary Fig. 1) during visual stimulation. These antagonists block NMDAR (*N*-methyl-D-aspartate receptors) and AMPAR (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptors), respectively. Therefore glutamate-receptor-dependent signalling is necessary for light-induced arbor elaboration.

Dendritic arbors grow through the selective stabilization of a fraction of newly added branches and the extension of those branches⁷⁻⁹. Visual stimulation significantly increased the number of new branches detected after 4 h (Fig. 1d, e; see also Supplementary Table 2). APV blocked the light-induced increase in branch additions and increased the rate of branch retractions (Fig. 1d; see also Supplementary Table 2). Newly added branches account for the increased branch length seen with light stimulation (Fig. 1e). Total branch length from branch additions and extensions is significantly greater when animals are exposed to light, compared with dark or compared with animals exposed to light in the presence of APV. The data indicate that visual stimulation enhances arbor growth by stabilizing newly added branches and suggest a role of NMDA receptors in activity-induced branch initiation and stabilization.

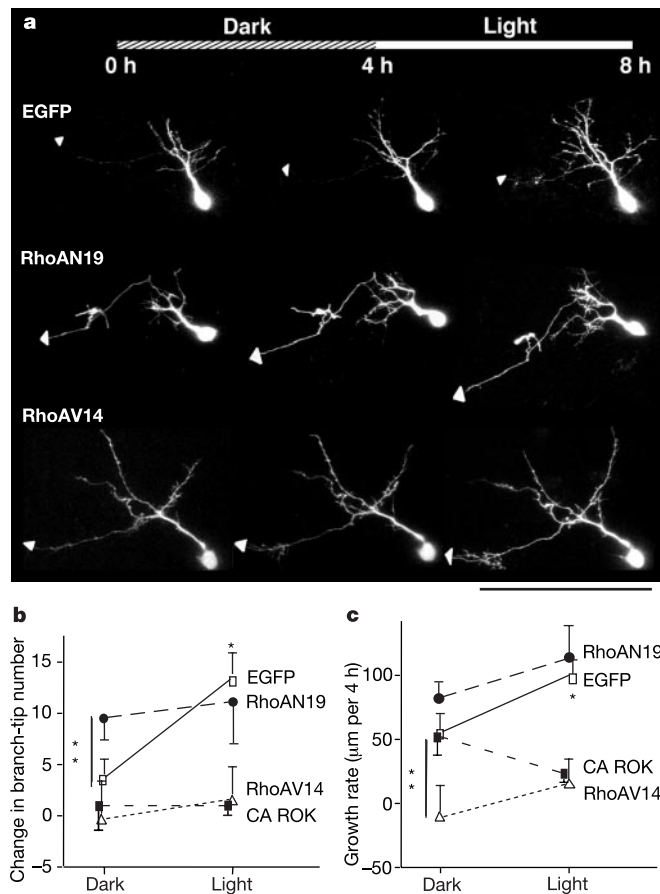


Figure 3 Decreased RhoA activity mediates light-induced dendritic arbor growth. **a**, Time-lapse *in vivo* confocal images collected at 4-h intervals over 8 h of tectal neurons expressing EGFP, dominant-negative RhoA (RhoAN19) or constitutively active (CA) RhoA (RhoAV14). Scale bar, 100 μm. **b**, **c**, Change in branch-tip number (**b**) and branch length (**c**) over 4 h in the dark (dark) or with visual stimulation (light). Asterisk, $P < 0.05$, Wilcoxon signed-rank test; double asterisk, $P < 0.05$, Mann-Whitney *U*-test.

The time course of the effect of visual stimulation on dendritic arbor dynamics was analysed by collecting two-photon microscope images of individual green fluorescent protein (GFP)-expressing tectal neurons at 1-h intervals over a period of 8 h (Fig. 2a). This stimulation and imaging protocol led to a significant light-induced increase in dendritic arbor growth rate and branch-tip number (Fig. 2d, g), as seen when animals were imaged three times over 8 h (Fig. 1). Despite the continuous addition and retraction of branches throughout 8 h (Fig. 2c, f), arbors showed a net increase in growth rate and branch-tip number during light stimulation, but not over

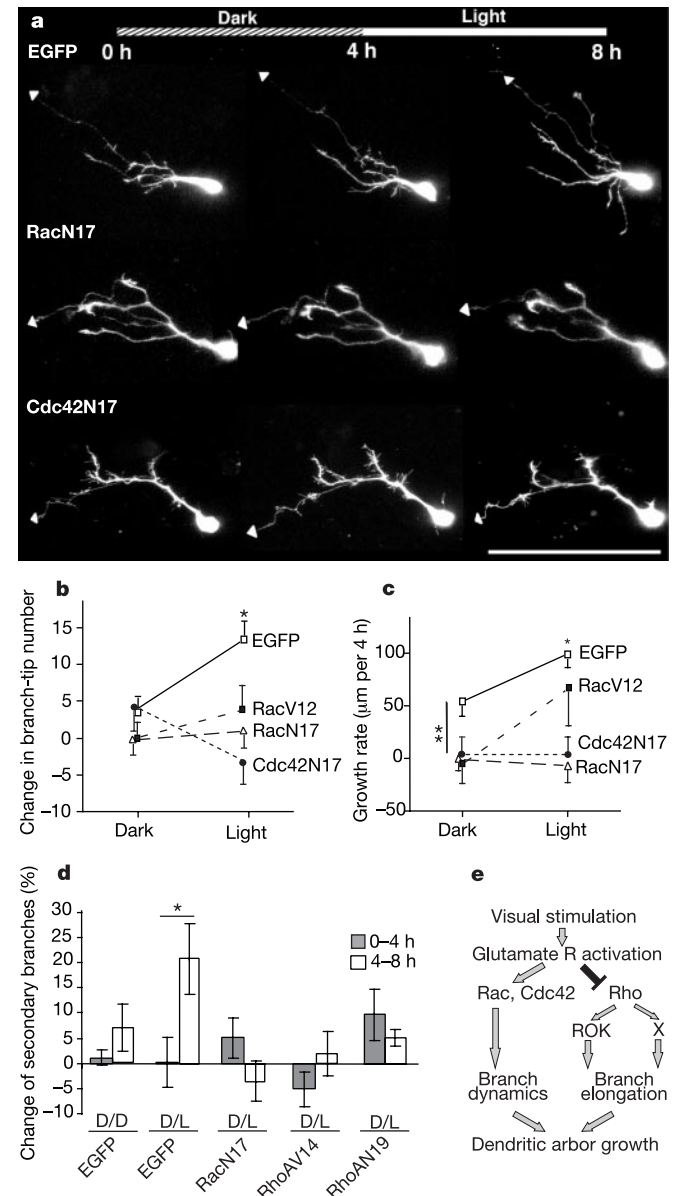


Figure 4 Rac and Cdc42 mediate the light-induced increase in dendrite branch number. **a**, Time-lapse *in vivo* confocal images of tectal neurons expressing EGFP, dominant-negative Rac (RacN17) or dominant-negative Cdc42 (Cdc42N17) over 8 h. Scale bar, 100 μm. **b**, **c**, Change in branch-tip number (**b**) and dendritic branch length (**c**) over 4 h in the dark (dark) or with visual stimulation (light). **d**, Change in secondary dendritic branches (expressed as per cent of initial total branch number) added to the primary dendrite during 4 h in the dark (D) or with visual stimulation (L) in neurons expressing the designated constructs. **e**, Summary of the role of visual stimulation and the Rho GTPases in dendritic arbor development (see text for details). Asterisk, $P < 0.05$, Wilcoxon signed-rank test; double asterisk, $P < 0.05$, Mann-Whitney *U*-test.

4 h in the dark (Fig. 2e, h). The change in branch length and branch-tip number is significantly greater by the second hour of visual stimulation than the change seen over 4 h in the dark. Five out of six neurons increased branch-length growth rates and branch-tip number in the first hour of visual stimulation compared with the previous hour in the dark (Fig. 2i, j). By contrast, most of the neurons decreased growth rates and branch dynamics during the first and second time windows in the dark (Fig. 2i, j). These data indicate that dendritic arbors respond within the first hour to visual stimulation with an increase in branch length and branch-tip number. The continued increase in branch-tip number and branch length with each hour of visual stimulation indicates that the system does not adapt to the visual stimulus over 4 h of visual stimulation.

Neuronal arbor elaboration requires the reorganization of the cytoskeleton. The Rho family of small GTPases regulates actin dynamics¹⁰ and has been implicated in the control of neuronal morphogenesis¹¹. Regulators and effectors of small GTPases such as kalirin-7, SynGAP and citron associate with NMDA receptors, providing potential biochemical links between NMDAR activity and GTPase-mediated regulation of cytoskeletal dynamics^{12–16}. However, no study has yet demonstrated a role for Rho GTPases under physiological conditions that control dendritic arbor development. Therefore we investigated the contribution of the Rho family of GTPases to the signalling mechanisms underlying activity-induced dendritic growth.

We have used recently an *in situ* binding assay to demonstrate that stimulation of the optic nerve significantly decreases endogenous RhoA activity and increases endogenous Rac activity in the optic tectum³. These changes require glutamate receptor activity. Here, we tested whether increased Rac and Cdc42 and decreased RhoA activity are required for light-induced dendritic arbor elaboration by collecting time-lapse images *in vivo* of optic tectal neurons co-expressing GFP and either constitutively active or dominant-negative forms of Rho GTPases. Constitutively active RhoAV14 blocked the light-induced dendritic arbor development that is seen in GFP-expressing neurons (Fig. 3a–c). These data indicate that RhoA activity inhibits activity-dependent dendritic arbor growth. They also suggest that visual stimulation promotes arbor growth by inhibiting RhoA, consistent with previous studies showing that decreased RhoA activity increases dendritic arbor elaboration⁴. If light stimulation activated a parallel pathway promoting dendritic arbor elaboration that is independent of RhoA activity, one would predict a synergistic effect of visual stimulation and expression of dominant-negative RhoAN19; however, light stimulation did not increase further dendritic arbor elaboration induced by dominant-negative RhoAN19 (Fig. 3a–c). This occlusion of light-induced dendritic arbor growth by dominant-negative RhoAN19 suggests that endogenous RhoA activity is relatively high in dendrites with slow growth rates and that decreasing RhoA activity increases dendritic arbor development. Therefore, in combination with our findings that optic nerve stimulation decreases endogenous RhoA activity³, these data suggest that RhoA is a downstream component of the light-induced pathway controlling dendritic arbor elaboration.

RhoA has been reported to negatively regulate dendrite outgrowth through ROK (also known as ROCK), a RhoA-binding kinase¹¹. Other potential RhoA effectors are mDia, protein kinase N, citron kinase, Lim kinase-1 and PRK2 (refs 17, 18). Expression of constitutively active ROK inhibited light-induced dendritic arbor growth (Fig. 3b, c). This suggests that the inhibition of activity-dependent dendritic arbor growth by RhoAV14 is at least partially mediated by enhanced ROK activity, consistent with the idea that light-induced arbor elaboration occurs when both RhoA and ROK activities are low.

Previous experiments⁴ and experiments included here indicate that Rac activity controls rates of branch additions and retractions.

Our *in situ* GTPase assays³ indicate that decreasing Rho activity increases Rac activity. This suggests that the increase in branch-tip number seen with inhibition of RhoA activity is probably due to activation of Rac. These data also suggest that regulation of branch dynamics is affected primarily by Rac activity, and that RhoA can alter Rac-mediated branch dynamics *in vivo*.

Tectal neurons expressing dominant-negative RacN17 and Cdc42N17 failed to respond to light stimulus with the increase in branch-tip number and branch length seen in GFP-expressing (EGFP) neurons (Fig. 4a–c; see also Supplementary Table 3). These data indicate that activation of Rac and Cdc42 is required for increased arbor elaboration in response to visual stimulation. Neurons expressing constitutively active RacV12 failed to show the significant light-induced increase in branch-tip number seen with EGFP controls (Fig. 4b, c). Although this is consistent with our observations that expression of constitutively active RacV12 increases RhoA activity³, which inhibits dendritic arbor growth, we cannot exclude the possibility that cycling of Rac from the GTP-bound to GDP-bound form may be required for its optimum activity¹⁹.

GTPase expression also affects arbor development in the absence of visual stimulation. Constitutively active RhoAV14 inhibited rates of branch additions and dendritic arbor growth in the dark compared with GFP-expressing control neurons ($P < 0.05$, Mann–Whitney *U*-test), whereas constitutively active ROK did not (Fig. 3b, c). Neurons expressing dominant-negative Cdc42N17 or dominant-negative RacN17 had significantly decreased rates of arbor growth and changes in branch-tip number compared with GFP-expressing control neurons ($P < 0.05$; Fig. 4a–c). These neurons could be imaged and appeared healthy for up to 10 days, indicating that expression of these mutants was not toxic. These observations are consistent with previous results^{4,11} and support the idea that the Rho GTPases are positioned at a bottleneck of multiple signal transduction pathways affecting neuronal structure¹¹.

Dendritic arbor development may proceed through repeated branching of dendritic growth cones or through a mechanism called ‘interstitial/back branching’ in which new branches emerge from more stable branches within the arbor²⁰. Tectal projection neurons imaged in this study have a single stable primary dendrite, an average of 12 secondary branches originating from the primary dendrite, and complex arborizations with up to fifth-order branches. Light stimulus significantly increased the number of secondary branches (Fig. 4d), indicating that interstitial branching is a prominent mechanism controlling dendritic arbor elaboration in response to visual stimulation *in vivo*. Expression of dominant-negative RacN17, Cdc42N17, RhoAN19 or constitutively active RhoAV14 prevented the light-induced increase in secondary branches. This indicates that the GTPases regulate the formation of interstitial branches in response to visual stimulation.

Our data indicate that visual system activity affects dendritic arbor elaboration through a mechanism involving Rho GTPases (Fig. 4e). Expression of mutant GTPases that interfere with endogenous GTPases blocks light-induced dendritic arbor development. We suggest that enhanced Rac and Cdc42 activity promotes branch dynamics. Decreased Rho activity promotes branch elongation mediated by several downstream effectors, including ROK. Rac- and Cdc42-mediated branch addition and stabilization and RhoA-mediated branch elongation⁴ cooperate to result in dendritic arbor growth.

Our data demonstrating that NMDAR activity and GTPases are both required for dendritic arbor development suggest that calcium influx through NMDAR affects regulators of Rho GTPase activity. Such regulation may occur through GTPase regulators or effectors such as SynGAP, kalarin or trio^{12,13,16,21}. The Rho GTPases may also affect intracellular calcium levels by regulating calcium influx into cells²² or calcium-induced calcium release from intracellular stores²³. Calcium-dependent events, including calcium- and calmo-

dulin-dependent kinase type II (CaMKII)-mediated branch stabilization²⁴, may cooperate with GTPase-mediated branch initiation and extension to control dendritic arbor development.

Glutamate-mediated synaptic activity is probably one of several regulators that affect dendritic arbor elaboration of tectal cells by activating the Rho GTPases. Several other extracellular signalling molecules that affect Rho GTPase activity include growth factors, adhesion or guidance molecules, cytokines, neurotransmitters and other neuroactive molecules²⁵. Indeed, the decreases in growth rates seen with expression of mutant GTPase constructs in the absence of visual stimulation probably represent the effects of these other signalling pathways on GTPase-mediated regulation of cytoskeletal structure.

Little is known about the mechanisms that link synaptic activity to structural changes in dendrites. Previous studies have indicated that actin-based spine dynamics are regulated by glutamate receptor activity⁶ and that afferent activity promotes the formation of dendritic structures^{2,4,5,7} and the stabilization of synaptic contacts^{2,26}. Here, we show a direct relation between sensory input, glutamate-mediated synaptic activity and GTPase-mediated dendritic structural plasticity in the intact animal (Fig. 4e). Our observations demonstrate that the initiation and selective stabilization of new dendritic branches occur in response to a visual stimulus. Our results indicate that visual stimulation affects structural plasticity *in vivo*, through NMDA receptor-induced intracellular signalling events mediated by the Rho GTPases. □

Methods

Image acquisition

Tadpoles were anaesthetized with 0.02% 3-aminobenzoic acid ethyl ester (MS222, Sigma) in Steinberg's solution for all procedures. Single tectal neurons were labelled by iontophoresis of DiI³ or by electroporation with pEGFP-C1 (Clontech), or with pEGFP-C1 and recombinant EGFP-GTPase plasmids in a 4:1 (GTPase:GFP) ratio²⁷. Confocal images were collected through a $\times 40$ Nikon oil immersion lens (1.3 NA) at 2- μ m steps through the entire z-dimension of labelled neurons⁵. Two-photon images were collected at 1.5- μ m steps with a custom-built microscope based on a modified Olympus Fluoview confocal scan box mounted on an Olympus BX50WI microscope with a Tsunami femtosecond-pulsed Ti:Sapphire laser. We used an Olympus LUM Plan F1/IR $\times 40$ water immersion objective (0.8 NA). Two-photon optical sections were an average of three frames. For the experiment in which neurons were imaged every hour over 8 h, animals were returned to a dark chamber in between image collection during the first 4-h period. During the second 4-h period animals were returned to the chamber with the visual stimulus in between imaging sessions. Tadpoles recovered from anaesthetic between imaging sessions.

Visual stimulus

Albino *Xenopus laevis* tadpoles were reared in ambient light until stage 46–47. Tadpoles have a functional visual system from stage 39/40 onward, in which retinal ganglion cell axons make glutamate-receptor-mediated synaptic connections with dendrites of tectal cells²⁸. Retinal ganglion cells respond well to a light on/light off stimulus. A 2-s step of light on or off evokes synaptic currents in tectal cells of *Xenopus* tadpoles²⁸. On the basis of these response properties, freely swimming tadpoles were placed individually in wells of a 12-well plate in a black Plexiglas chamber with a 3×4 panel of green light-emitting diodes (LEDs; λ_{max} 567 nm, AND191GCP; Allied Electronics) on the top of the chamber. Each row of LEDs turned on and remained on for 1 s at a frequency of 0.2 Hz. The rows turned on and off sequentially to create a simulated motion stimulus, with 1 s of darkness between each cycle. *Xenopus* tadpole tectal cells can respond to repeated stimulation of the same region of retina without apparent adaptation²⁸. In our experiments, adaptation of tectal cells to the stimulus is not likely to be significant because the animals are swimming freely during the stimulation period, so the apparent location and direction of the stimulus vary constantly. Light exposure during the 4-h period in the absence of visual stimulation was kept to a minimum, but animals were exposed to some light during imaging. Some animals were exposed to 100 μ M DL-APV or 20 μ M CNQX during visual stimulation. MK801, another NMDA receptor antagonist, had a similar inhibitory effect on light-induced growth to that seen with APV (relative growth rate: 0.94 ± 0.73 , $n = 9$). Treatment of tectal cells with APV while the animals were in the dark for 4 h did not significantly affect dendritic growth (relative growth rate: 1.34 ± 0.5 , $n = 9$, $P = 0.7$).

Image analysis

We used a 4-h imaging protocol on the basis of our previous studies which showed that 4 h is sufficiently long to detect quantifiable increases in dendritic arbor branch length and branch-tip number, but short enough so that images of arbors from sequential time points can be reliably superimposed⁹. We measured the total dendritic branch length (TDBL) and branch-tip number of the dendritic arbors by reconstructing the whole dendritic arbor in

three dimensions using Object Image (<http://simon.bio.uva.nl/object-image.html>). The software computed the TDBL and branch-tip number using custom macros written by E.S.R.²⁹. Values are expressed as mean and standard error. Statistical analysis was done with Mann-Whitney *U* nonparametric statistical analysis and Wilcoxon signed-rank test. Analysis was performed blind to the type of treatment. Between 9 and 24 neurons per treatment were analysed, unless stated otherwise.

Construction of expression plasmids of GTPase mutants

Fragments (600 base pairs) containing the full coding sequence of RacN17, Cdc42N17, RhoAN19 and RhoV14 were cloned in-frame into pEGFP-C1 (Clontech Laboratories) and expressed as GFP-fusion proteins³⁰.

Received 10 April; accepted 25 June 2002; doi:10.1038/nature00987.

1. Builluart, P. et al. Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392**, 823–826 (1998).
2. Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
3. Li, Z., Aizenman, C. D. & Cline, H. T. Regulation of Rho GTPases by crosstalk and neuronal activity *in vivo*. *Neuron* **33**, 741–750 (2002).
4. Li, Z., Van Aelst, L. & Cline, H. T. Rho GTPases regulate distinct aspects of dendritic arbor growth in *Xenopus* central neurons *in vivo*. *Nature Neurosci.* **3**, 217–225 (2000).
5. Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
6. Fischer, M., Kaech, S., Wagner, U., Brinkhaus, H. & Matus, A. Glutamate receptors regulate actin-based plasticity in dendritic spines. *Nature Neurosci.* **3**, 887–894 (2000).
7. Wong, W. T., Faulkner-Jones, B. E., Sanes, J. R. & Wong, R. O. Rapid dendritic remodeling in the developing retina: dependence on neurotransmission and reciprocal regulation by Rac and Rho. *J. Neurosci.* **20**, 5024–5036 (2000).
8. Wu, G.-Y., Zou, D. J., Rajan, I. & Cline, H. T. Dendritic dynamics *in vivo* change during neuronal maturation. *J. Neurosci.* **19**, 4472–4483 (1999).
9. Rajan, I. & Cline, H. T. Glutamate receptor activity is required for normal development of tectal cell dendrites *in vivo*. *J. Neurosci.* **18**, 7836–7846 (1998).
10. Van Aelst, L. & D'Souza-Schorey, C. Rho GTPases and signaling networks. *Genes Dev.* **11**, 2295–2322 (1997).
11. Luo, L. Rho GTPases in neuronal morphogenesis. *Nature Rev. Neurosci.* **1**, 173–180 (2000).
12. Penzes, P. et al. The neuronal Rho-GEF Kalirin-7 interacts with PDZ domain-containing proteins and regulates dendritic morphogenesis. *Neuron* **29**, 229–242 (2001).
13. Kim, J. H., Liao, D., Lau, L.-F. & Hagan, R. L. SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90 protein family. *Neuron* **20**, 683–691 (1998).
14. Zhang, W., Vazquez, L., Apperson, M. & Kennedy, M. B. Citron binds to PSD-95 at glutamatergic synapses on inhibitory neurons in the hippocampus. *J. Neurosci.* **19**, 96–108 (1999).
15. Furuyashiki, T. et al. Citron, a Rho-target, interacts with PSD-95/SAP-90 at glutamatergic synapses in the thalamus. *J. Neurosci.* **19**, 109–118 (1999).
16. Chen, H. J., Rojas, S. M., Oguni, A. & Kennedy, M. B. A synaptic Ras-GTPase activating protein (p135 SynGAP) inhibited by CaM kinase II. *Neuron* **20**, 895–904 (1998).
17. Lawler, S. Regulation of actin dynamics: the LIM kinase connection. *Curr. Biol.* **9**, R800–R802 (1999).
18. Aspenstrom, P. Effectors for the Rho GTPases. *Curr. Opin. Cell Biol.* **11**, 95–102 (1999).
19. Albertinazzi, C., Gilardelli, D., Paris, S., Longhi, R. & de Curtis, I. Overexpression of a neural-specific rho family GTPase, cRac1B, selectively induces enhanced neurite outgrowth and neurite branching in primary neurons. *J. Cell Biol.* **142**, 815–825 (1998).
20. Acebes, A. & Ferrus, A. Cellular and molecular features of axon collaterals and dendrites. *Trends Neurosci.* **23**, 557–565 (2000).
21. Debant, A. et al. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc. Natl Acad. Sci. USA* **93**, 5466–5471 (1996).
22. Wilk-Blaszczak, M. et al. The monomeric G-proteins Rac1 and/or Cdc42 are required for the inhibition of voltage-dependent calcium current by bradykinin. *J. Neurosci.* **17**, 4094–4100 (1997).
23. Djoudier, N., Prepens, U., Aktories, K. & Cavaliere, A. Inhibition of calcium release-activated calcium current by Rac/Cdc42-inactivating clostridial cytotoxins in RBL cells. *J. Biol. Chem.* **275**, 18732–18738 (2000).
24. Wu, G.-Y. & Cline, H. T. Stabilization of dendritic arbor structure *in vivo* by CaMKII. *Science* **279**, 222–226 (1998).
25. Kjoller, L. & Hall, A. Signaling to Rho GTPases. *Exp. Cell Res.* **253**, 166–179 (1999).
26. Lendvai, B., Stern, E. A., Chen, B. & Svoboda, K. Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex *in vivo*. *Nature* **404**, 876–881 (2000).
27. Haas, K., Sin, W.-C., Javaherian, A., Li, Z. & Cline, H. T. Single-cell electroporation for *in vivo* neuronal gene expression. *Neuron* **29**, 1–9 (2001).
28. Zhang, L. I., Tao, H. W. & Poo, M. Visual input induces long-term potentiation of developing retinotectal synapses. *Nature Neurosci.* **3**, 708–715 (2000).
29. Ruthazer, E. S. & Cline, H. T. Multiphoton imaging of neurons in living tissue: acquisition and analysis of time-lapse morphological data. *J. Real-Time Imag.* **8**, 175–188 (2002).
30. Leung, T., Chen, X. Q., Manser, E. & Lim, L. The p160 RhoA-binding kinase ROK α is a member of a kinase family and is involved in the reorganization of the cytoskeleton. *Mol. Cell. Biol.* **16**, 5313–5327 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank K. Svoboda, P. O'Brien and B. Burbach for help constructing the two-photon microscope and P. O'Brien for constructing the electronic circuit for visual stimulation. We also thank K. Bronson for technical assistance, J. Duffy for assistance with the figures, and T. Leung and J. Dong for cDNA. We are grateful to L. Van Aelst for discussions, and

R. Malinow, J. Dubnau and members of the Cline laboratory for comments on the manuscript. This work was supported by the NIH (H.T.C., K.H., E.S.R.) and an endowment from the Charles Robertson Foundation to H.T.C..

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.C. (e-mail: cline@cshl.org).

ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum

Thomas Serwold, Federico Gonzalez, Jennifer Kim, Richard Jacob & Nilabh Shastri

Division of Immunology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3200, USA

The ability of killer T cells carrying the CD8 antigen to detect tumours or intracellular pathogens requires an extensive display of antigenic peptides by major histocompatibility complex (MHC) class I molecules on the surface of potential target cells¹. These peptides are derived from almost all intracellular proteins and reveal the presence of foreign pathogens and mutations. How cells produce thousands of distinct peptides cleaved to the precise lengths required for binding different MHC class I molecules remains unknown^{2,3}. The peptides are cleaved from endogenously synthesized proteins by the proteasome in the cytoplasm^{4,5} and then trimmed by an unknown aminopeptidase in the endoplasmic reticulum (ER)^{6–8}. Here we identify ERAAP, the aminopeptidase associated with antigen processing in the ER. ERAAP has a broad substrate specificity, and its expression is strongly upregulated by interferon- γ . Reducing the expression of ERAAP through RNA interference prevents the trimming of peptides for MHC class I molecules in the ER and greatly reduces the expression of MHC class I molecules on the cell surface. Thus, ERAAP is the missing link between the products of cytosolic processing and the final peptides presented by MHC class I molecules on the cell surface.

To identify the ER aminopeptidase, we used detergent to solubilize microsomes from mouse liver and spleens. Ion exchange chromatography of solubilized microsomes showed a predominant peak of activity that was blocked by the aminopeptidase inhibitor leucinethiol (Fig. 1a). This activity was purified by several chromatographic steps (Methods). The highly enriched fractions from the final purification step contained one main band with a relative molecular mass (M_r) of roughly 100,000 (100K) on a SDS polyacrylamide gel electrophoresis (SDS–PAGE) gel stained with Coomassie blue (Fig. 1b). We subjected this band to in-gel trypsin digestion followed by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry and used the resulting tryptic peptide fingerprint to search the National Center for Biotechnology Information database^{9–11}.

In two independent experiments, 25 masses matched the predicted output of trypsin-digested murine adipocyte-derived leucine aminopeptidase (accession code AF227511)¹². This assignment was verified with the amino acid sequences of four peptides determined by high-energy collision-induced dissociation tandem mass spec-

trometry analysis (Fig. 1c)¹³. This murine aminopeptidase of 930 amino acids and its probable human and rat orthologues are also known as puromycin-insensitive leucyl-specific aminopeptidase^{12,14,15}. On the basis of our results, we hereafter refer to this enzyme as ERAAP (for ER aminopeptidase associated with antigen processing) to designate its intracellular location and function in trimming peptides in the MHC class I antigen processing pathway.

ERAAP is a member of the M1 family of zinc metalloproteases that are defined by a highly conserved His-Glu-X-X-His-X₁₈-Glu motif in the core peptidase unit (ref. 16 and Fig. 1c). The amino terminus contains a hydrophobic leader sequence, which indicated that it might be cotranslationally translocated into the ER. Hydrophobicity plots did not predict any other transmembrane domains, and there were no other obvious motifs to predict its intracellular location. To determine its location, therefore, we first transfected COS cells with the murine ERAAP complementary DNA and carried out western blot analysis with the whole-cell lysates and the culture supernatants. A single band close to the predicted

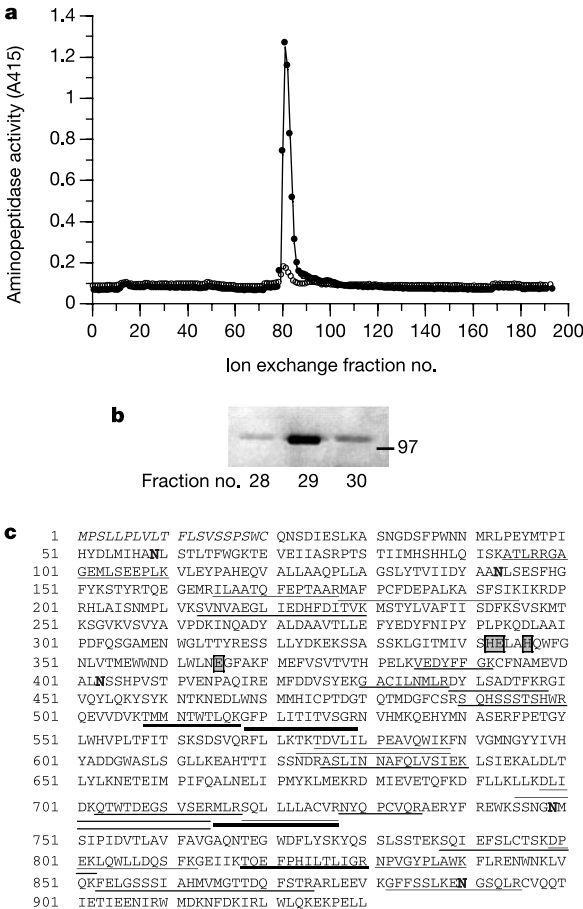


Figure 1 Purification and identification of ERAAP. **a**, Microsomes from mouse livers and spleens were lysed in detergent and separated by ion exchange chromatography. Fractions were tested for aminopeptidase activity by incubating with leucine *p* nitroanilide without (filled circles) or with (open circles) the aminopeptidase inhibitor leucinethiol. **b**, SDS–PAGE gel stained with Coomassie blue of three consecutive fractions containing aminopeptidase activity from the final hydrophobic interaction chromatography purification step. **c**, The sequence of murine ERAAP. Underlined peptides matched the masses identified by mass spectrometry. Peptides identified by sequencing are underlined in bold. The ER translocation signal at the N terminus is in italics, active-site residues are boxed, and putative *N*-linked glycosylation sites are in bold.

molecular weight of 106K was detected mainly in the cell pellet and only faintly in the culture supernatant, showing that the bulk of ERAAP was intracellular (Fig. 2a).

Next, we labelled EL4 cells briefly with [35 S]methionine, chased them for the indicated times, and immunoprecipitated them with an antiserum to ERAAP. The immunoprecipitates were treated with endoglycosidase H (EndoH), which removes the high-mannose oligosaccharides from ER proteins but cannot cleave the modified oligosaccharides of proteins that have traversed the Golgi. The 100K ERAAP band was detected at the beginning of the chase and was sensitive to EndoH, as indicated by the decrease in its apparent M_r (Fig. 2b). ERAAP was sensitive to EndoH over the whole 8-h period of chase, which indicated that it was retained in an early secretory compartment.

To visualize the intracellular location of ERAAP, we stained COS cells expressing murine ERAAP with antiserum to ERAAP and analysed them by immunofluorescence microscopy. A reticular, perinuclear staining pattern characteristic of the ER was observed specifically in cells transfected with ERAAP but not in cells transfected with control vector (Fig. 2c). In addition, ERAAP colocalized mainly with the ER markers BiP and gp96. Thus, ERAAP is localized in the ER.

MHC class I molecules are expressed by almost all tissues. To determine whether ERAAP expression correlated with that of MHC class I molecules, we analysed seven different mouse tissues by western blot for expression of ERAAP, MHC class I molecules and the housekeeping proteins actin and gp96 as controls. ERAAP was

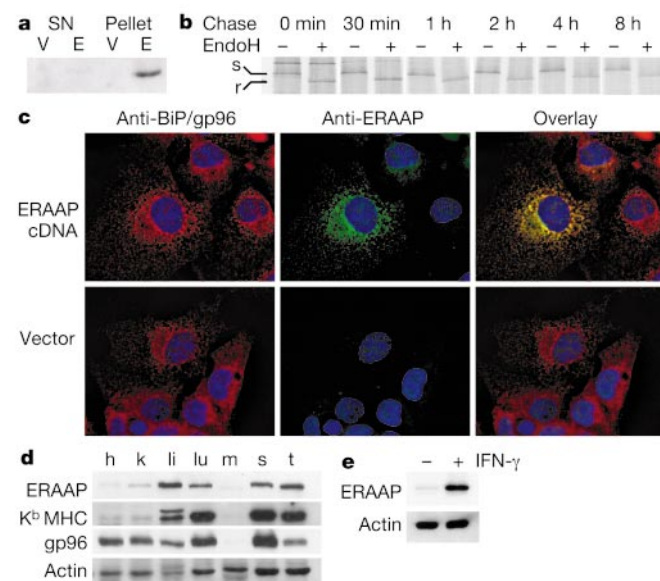


Figure 2 ERAAP is located in the ER and its tissue distribution correlates with that of MHC class I molecules. **a**, COS cells were transfected with murine ERAAP (E) or vector control (V) and their supernatants (SN), and cell pellets were analysed by western blot using rabbit antibodies against mouse ERAAP. **b**, EL4 cells were labelled with [35 S]methionine and chased with cold methionine. At the indicated time points, cells were lysed and ERAAP was immunoprecipitated. Aliquots were either left untreated (–) or digested with EndoH (+) before SDS–PAGE and autoradiography. The EndoH sensitive (s) and resistant (r) forms of ERAAP are indicated. **c**, COS cells transfected with murine ERAAP or vector control were stained with antisera to ERAAP (green) and antibodies to BiP and gp96 (red). Nuclei are stained blue. An overlay of the green and red stains is shown on the right in yellow. **d**, Western blot of tissues for ERAAP, K^b MHC class I molecules, gp96 and actin. h, heart; k, kidney; li, liver; lu, lung; m, muscle; s, spleen; t, thymus. **e**, B6 tail fibroblast cells were either left untreated or treated with 100 units ml^{–1} IFN-γ for 24 h. Western blots were carried out on whole-cell lysates and reprobed for actin as a loading control.

detected in all tissues analysed, but its expression was highest in liver, lung, spleen and thymus—the tissues that expressed the highest concentrations of MHC class I molecules (Fig. 2d). Expression of ERAAP was also increased about tenfold by treating cultured fibroblasts with the inflammatory cytokine interferon-γ (IFN-γ), which also upregulates many other key components in the antigen processing pathway (ref. 17 and Fig. 2e).

To assess the importance of ERAAP in generating peptides for presentation by MHC class I molecules, we reduced the expression of ERAAP with small interfering RNA oligonucleotide duplexes (siRNA)¹⁸. L^d-L cells were transfected with siRNA specific for either mouse ERAAP or, as a negative control, human ERAAP, which differs by eight nucleotides in the targeted region. Transfection of

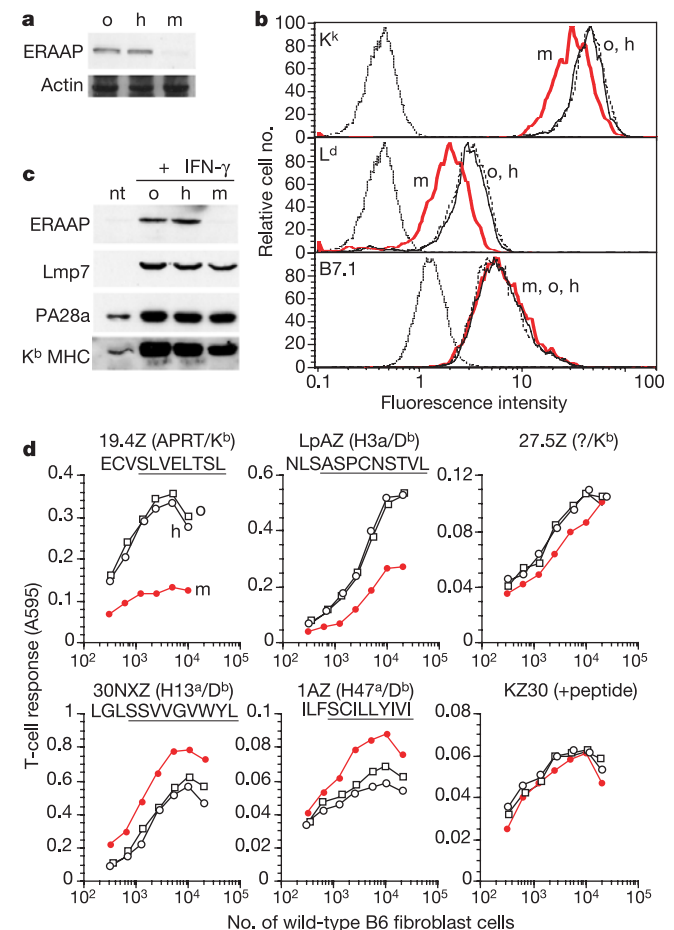


Figure 3 Reduction in ERAAP expression by RNAi affects antigen presentation. **a**, Mouse L^d-L cells were transfected with transfection reagent alone (o), or siRNA specifically targeting mouse ERAAP (m) or human ERAAP (h). After 72 h, cells were tested for ERAAP and actin expression by western blot. **b**, Flow cytometry profiles of L^d-L cells stained for surface expression of K^k and L^d MHC class I molecules and the negative control B7.1. Dotted lines represent secondary antibody alone, broken and unbroken lines represent oligofectamine (o) and human siRNA (h), respectively. Bold red lines represent mouse siRNA (m). **c**, Tail fibroblasts from B6 mice were transfected as in **a** and after 2 d treated with IFN-γ for 24 h, and then analysed by western blot for ERAAP, Lmp7, PA28α and K^b MHC class I molecules. **d**, Cells transfected with oligofectamine (o), human ERAAP (h), or mouse ERAAP (m) in **c** were co-cultured with the indicated T-cell hybridomas for 12 h. T-cell activation was assayed by β-galactosidase response. The peptide–MHC specificity of each T cell is indicated, as well as the identity of each peptide (underlined) and a few N-terminal flanking residues. As a control for toxicity and APC number, the response of the KZ30 T cell to a 10 nM concentration of its cognate peptide was also determined.

cells with siRNA complementary to mouse ERAAP reduced the amount of ERAAP by about 90%, whereas transfection with the siRNA targeting human ERAAP had no effect (Fig. 3a). In cells with reduced ERAAP expression, the surface quantities of the MHC class I molecules K^k and L^d decreased by 23% and 44%, respectively, but there was no change in the expression of B7.1 (Fig. 3b). Because empty MHC class I molecules are not expressed stably on the cell surface, this result implicates ERAAP in peptide supply.

We tested the role of ERAAP in generating specific peptide–MHC complexes in antigen presentation assays. C57BL/6 (B6) mouse fibroblast cells were transfected with siRNA directed against mouse or human ERAAP and were treated with IFN- γ to induce upregulation of the antigen processing pathway and expression of MHC class I molecules. Western blots showed that the transfected cells responded to IFN- γ by upregulating MHC class I molecules, and also Lmp7 and PA28 α —two proteins involved in antigen processing^{19,20}. Notably, cells transfected with mouse siRNA, but not control siRNA, failed to upregulate ERAAP specifically (Fig. 3c).

We used the B6 tail fibroblast cells that had been treated with

siRNA as antigen-presenting cells (APCs) to assay for T cells specific for different endogenous peptides presented by K^b or D^b MHC class I molecules (ref. 6 and Fig. 3d). The expression of two peptides presented by K^b and D^b and recognized by bm19.4Z and LpAZ T cells, respectively, was inhibited strongly in cells transfected with mouse ERAAP siRNA as compared with cells treated with transfection reagent or human siRNA. Presentation of another yet unknown peptide presented by K^b to the 27.5Z T cell was slightly diminished, whereas that of two other peptides presented by D^b to 30NXZ and 1AZ T cells was enhanced reproducibly. These results concur with those of our previous study in which treating cells with the aminopeptidase inhibitor leucinethiol caused a similar inhibition and enhancement of the same peptides⁶. Thus, ERAAP is involved in generating a large fraction of antigenic peptides. In the absence of ERAAP the concentrations of some peptides increase, possibly because ERAAP normally destroys those peptides or because there is decreased competition for binding MHC class I molecules from ERAAP-dependent peptides.

To confirm that ERAAP trims peptides in the ER, we generated fibroblast cells deficient in transporter associated with antigen processing (TAP) that could not import peptides from the cytoplasm into the ER. We targeted precursors directly into the ER of these cells by an ER translocation sequence (ES) to allow analysis of peptide processing exclusively in the ER (refs 6, 21, and Fig. 4a). Cells expressed one of two different precursors of the SHL8 peptide: one that directly yielded SHL8 after signal peptidase cleavage (ES–SHL8), and one that required the removal of seven N-terminal flanking residues to yield SHL8 (ES–[X7]SHL8). Reduction of ERAAP expression specifically inhibited generation of the SHL8 peptide from the N-terminally extended precursor (Fig. 4b), as measured by the response of SHL8– K^b -specific B3Z T cells. In contrast, reduced ERAAP expression had no effect on generating SHL8 from the precursor that required no trimming (Fig. 4a). Thus, ERAAP is required specifically for trimming extended precursors in the ER.

To determine its substrate specificity, we incubated purified ERAAP with SHL8 peptide that had been extended by various amino acids at the N terminus. After incubation with the decapeptide substrates for 0, 30 or 90 min, the peptide digests were separated by high-performance liquid chromatography (HPLC), and the products were detected by the B3Z T cell after all peptides had been converted to SHL8 with trypsin (ref. 22 and Fig. 4c). ERAAP cleaved N-terminal lysine (K), leucine (L), tyrosine (Y) and asparagine (N) residues to yield nonamer and octamer products in 30 min. At 90 min the SHL8 octapeptide was the predominant product, although smaller products cannot be detected in this T-cell-based assay²³. By contrast, ERAAP failed to cleave the N-terminal lysine residue when it was followed by proline. This inability to cleave the ‘X-Pro’ bond has been established as a key characteristic of the ER aminopeptidase⁶, and therefore strongly implicates ERAAP as the aminopeptidase responsible for providing peptides for the many MHC class I molecules that present peptides with the ‘X-Pro-X_n’ motif. The otherwise broad specificity of ERAAP suits its role in cleaving a wide spectrum of peptides for MHC class I molecules.

Studies have shown that proteolysis of antigenic precursors in the cytoplasm rarely yields the final peptide^{5,24,25}. The discovery of ERAAP and its role in trimming peptides in the ER now fills a central gap, which was recognized a decade ago², in our understanding of antigen processing. Cytoplasmic proteolysis generates only the carboxy terminus of antigenic peptides²⁴. These N-terminally extended peptides, after transport by TAP, have the opportunity to bind MHC class I molecules in the ER. If the peptides are too long, then their N termini are trimmed by ERAAP until they acquire the optimal length; the peptide–MHC complex can then leave the ER. This model can now explain why peptides presented by MHC class I possess precise lengths.

To our knowledge, ERAAP is the only luminal protease with a

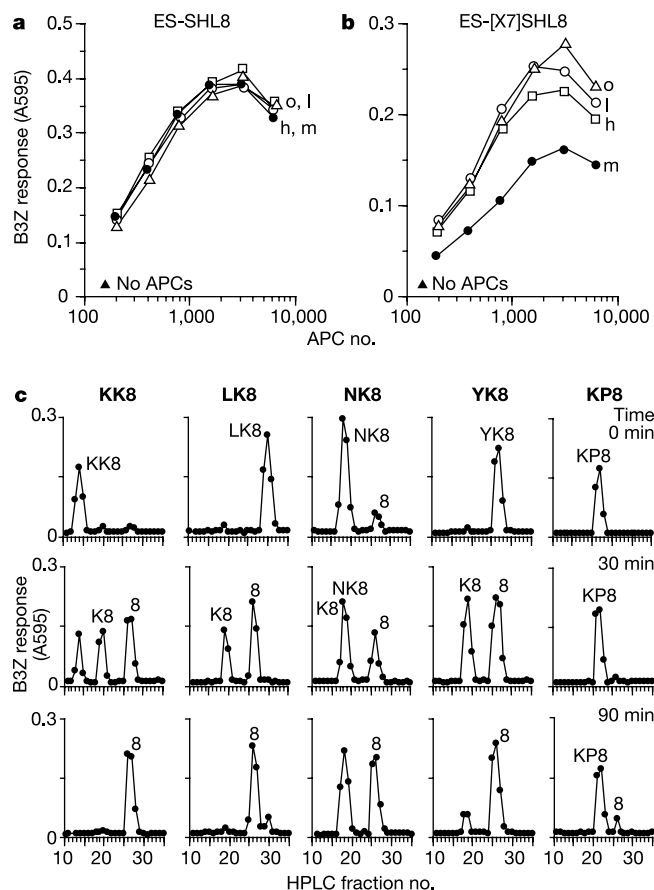


Figure 4 ERAAP is required for trimming peptides in the ER. **a, b**, TAP-knockout fibroblasts expressing either ES-SHL8 (**a**) or ES-[X7]SHL8 precursors (**b**) were transfected with oligofectamine alone (open triangles), siRNA specific for lamin b1 (open circles)¹⁸ or human ERAAP (squares) as negative controls, or mouse ERAAP (filled circles). Generation of the SHL8 peptide was measured by adding B3Z T cells specific for the SHL8– K^b complex as in Fig. 3d. **c**, Substrate specificity of ERAAP was determined by incubating purified ERAAP with SHL8 peptide analogues with the N-terminal amino acids indicated above each graph. Peptide digestion products were analysed at 0 (top), 30 (middle) and 90 (bottom) min, by assaying the HPLC fractionated mixture with T cells specific for SHL8– K^b . HPLC fractions were treated with trypsin to liberate SHL8, which enabled the detection of all N-terminally extended peptides²². Amino acids are shown in single-letter code and ‘8’ refers to the SHL8 octapeptide.

broad substrate specificity in the ER. The presence of constitutively active ERAAP in the ER suggests that there are mechanisms to prevent its potentially destructive cleavage of nascent polypeptides as they fold in this compartment. Possibly, ERAAP might be sequestered in a location that is inaccessible to these large substrates, or the specificity of ERAAP might be restricted only to oligopeptides or peptides whose C termini are already bound by MHC class I molecules⁷. Manipulation of ERAAP function provides an attractive method to modulate the responses of T cells expressing the CD8 antigen. By contrast to the proteasome, if ERAAP is not involved in essential cellular processes then it might represent a specific therapeutic target for manipulating the MHC class I antigen processing pathway. □

Methods

Purification and identification of ERAAP

We purified ERAAP from detergent-solubilized liver and spleen microsomes of 50 mice⁷. ERAAP was precipitated with 2.27 M (NH₄)₂SO₄ after pre-clearing material that was insoluble in 1.9 M (NH₄)₂SO₄. The pellet was resuspended in 10 mM NaHPO₄ plus 1% CHAPS buffer and was fractionated by size using the same buffer over two G3000SWXL columns (Tosohas) in series. Aminopeptidase-containing fractions were fractionated by ion exchange using a DEAE 301VHP575 column (Vydac) with a NaCl gradient in 10 mM Tris plus 0.1% CHAPS buffer (pH 7.5). The peak of aminopeptidase activity eluted at roughly 50 mM NaCl and was separated further by hydrophobic interaction chromatography on a phenyl-5PW column (Tosohas) using a gradient from 1.8 M to 0 M (NH₄)₂SO₄ in 10 mM Tris (pH 7.5). The main peak of aminopeptidase activity eluted at 210 mM (NH₄)₂SO₄ (Fig. 1b). These fractions were concentrated by centrifugation through a 10K cut-off Ultrafree MC filter (Millipore) and analysed by SDS-PAGE. We monitored aminopeptidase activity by adding an aliquot of the fractions to 2 mg ml⁻¹ leucine p-nitroanilide in PBS plus 0.1% Triton X-100. After a 30-min incubation at 37 °C, substrate cleavage was measured by the change in absorbance at 415 nm. The aminopeptidase inhibitor leucinethiol was used at a final concentration of 60 μM.

Transfections and immunoprecipitations

We transfected COS cells using Eugene 6 (Roche) according to the manufacturer's instructions with either pcDNA1 (Invitrogen) or pcDNA1 containing murine ERAAP cDNA cloned by polymerase chain reaction with reverse transcription. The cells were collected after 48 h. Equal fractions of cells (2 × 10⁵) and supernatant (10 μl) were separated by SDS-PAGE and blotted onto 0.45-μm nitrocellulose membranes (Amersham).

We carried out western blots with 2 μg ml⁻¹ affinity-purified rabbit antiserum generated against the N terminus of the ERAAP peptide, QNSDIESLKASNGD, and a donkey peroxidase-conjugated secondary antibody against rabbit IgG (Amersham). Blots were developed using Supersignal West Femto substrate (Pierce) and exposed to film for 2–60 s. The other antibodies were rabbit antiserum to the C-terminal tail of K^b (ref. 7), goat antibody to actin (Santa Cruz Biotechnology), mouse antibody to BiP/gp96 (StressGen), and rabbit antibody to PA28α and rabbit antibody to Lmp7 (Affinity Bioreagents).

We starved EL4 cells (3.5 × 10⁷) of methionine for 10 min and then labelled them for 15 min with 100 μCi ml⁻¹ EasyTag ³⁵S-labelling mix (Dupont NEN) in medium plus 2% dialysed fetal calf serum. The cells were then placed in fresh medium containing 10 mM unlabelled methionine. At the indicated time points, 8 × 10⁶ cells were removed and lysed for 30 min on ice in 1 ml of PBS plus 1% Triton X-100 containing protease inhibitors. The solubilized material was pre-cleared with Staph A and Staph G (Zymed Laboratories) plus rabbit serum. Protein A plus protein G beads (Zymed) pre-incubated with rabbit antibody to ERAAP were added to the lysate and rocked for 2 h. Immunoprecipitated material was washed extensively, resuspended in SDS sample buffer and divided into two equal portions. One portion was treated for 1 h at 37 °C with 500 units of EndoH (New England Biolabs); the other was left untreated. We separated the samples by SDS-PAGE and analysed them by autoradiography. COS cells were transfected and analysed by immunofluorescence staining as described²⁶.

RNA interference

The coding strands of the RNA oligonucleotide duplexes directed to the 5' end of mouse or human ERAAP messenger RNA were 5'-AGCUAGUAAUGGAGACUCA-3' and 5'-CGUAGUGAUGGACACCAU-3', respectively (Dharmacon). The oligonucleotide for lamin b1 has been described¹⁸. All oligonucleotides had 3' dTdT overhangs. RNA duplexes were annealed as recommended by the manufacturer. L^d-L cells or tail fibroblast cell lines from B6 wild-type or B6 TAP knockout mice were grown in 24-well dishes, and oligonucleotides were transfected into cells using oligofectamine (Invitrogen) according to the manufacturer's instructions. We assayed cells by western blot, flow cytometry and/or antigen presentation 72 h after transfection. For flow cytometry, monoclonal antibodies 30-5-7S (anti-L^d) and Y3 (anti-K^b) were used with fluorescein isothiocyanate (FITC)-conjugated goat antibody to mouse IgG. Flow cytometry was carried out on a FacScan (Coulter). For antigen presentation assays, cells were titrated in 96-well plates, and the indicated LacZ-inducible T-cell hybridomas were added (10⁵ per well). We monitored T-cell activation by the accumulation of LacZ using chlorophenolred-β-D-galactopyranoside as a substrate⁶.

Enzyme assays

We purified ERAAP from mouse microsomes through the ion-exchange step described above. Each reaction contained 50 μl of ERAAP (equivalent to 0.12 nmol LPNA cleaving activity per μl per min) and 3 pmol of peptide substrate in PBS plus 0.1% Triton X-100. The reactions were stopped at the indicated times by adding 400 μl of 10% acetic acid. We separated the peptide mixture by reversed-phase HPLC and assayed the fractions after trypsin treatment to allow the detection of N-terminally extended analogues²².

Received 12 June; accepted 31 July 2002; doi:10.1038/nature01074.

- Shastri, N., Schwab, S. & Serwold, T. Producing nature's gene-chips. The generation of peptides for display by MHC class I molecules. *Annu. Rev. Immunol.* **20**, 463–493 (2002).
- Falk, K., Rötzschke, O. & Rammensee, H. G. Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature* **348**, 248–251 (1990).
- Falk, K., Rötzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290–296 (1991).
- Rock, K. L. & Goldberg, A. L. Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **17**, 739–779 (1999).
- Kloetzel, P.-M. Antigen processing by the proteasome. *Nature Rev. Mol. Cell Biol.* **2**, 179–187 (2001).
- Serwold, T., Gaw, S. & Shastri, N. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nature Immunol.* **2**, 644–651 (2001).
- Brouwenstijn, N., Serwold, T. & Shastri, N. MHC class I molecules can direct proteolytic cleavage of antigenic precursors in the endoplasmic reticulum. *Immunity* **15**, 95–104 (2001).
- Fruci, D., Niedermann, G., Butler, R. H. & van Ender, P. M. Efficient MHC Class I-independent amino-terminal trimming of epitope precursor peptides in the endoplasmic reticulum. *Immunity* **15**, 467–476 (2001).
- Hellman, U., Wernstedt, C., Genez, J. & Heldin, C. H. Improvement of an 'In-Gel' digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* **224**, 451–455 (1995).
- Clauser, K. R., Baker, P. & Burlingame, A. L. Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871–2882 (1999).
- Helen, I. & Field, E. D. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics* **2**, 36–47 (2001).
- Hattori, A., Matsumoto, H., Mizutani, S. & Tsujimoto, M. Molecular cloning of adipocyte-derived leucine aminopeptidase highly related to placental leucine aminopeptidase/oxytocinase. *J. Biochem. (Tokyo)* **125**, 931–938 (1999).
- Medzhradszky, K. F. E. et al. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552–558 (2000).
- Schomburg, L., Kollmus, H., Friedrichsen, S. & Bauer, K. Molecular characterization of a puromycin-insensitive leucyl-specific aminopeptidase, PILS-AP. *Eur. J. Biochem.* **267**, 3198–3207 (2000).
- Miyashita, H. et al. A mouse orthologue of puromycin-insensitive leucyl-specific aminopeptidase is expressed in endothelial cells and plays an important role in angiogenesis. *Blood* **99**, 3241–3249 (2002).
- Rawlings, N. D., O'Brien, E. A. & Barrett, A. J. MEROPS: the protease database. *Nucleic Acids Res.* **30**, 343–346 (2002).
- Boehm, U., Klamp, T., Groot, M. & Howard, J. C. Cellular responses to interferon-γ. *Annu. Rev. Immunol.* **15**, 749–795 (1997).
- Elbashir, S. M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
- Fehling, H. J. et al. MHC class I expression in mice lacking the proteasome subunit LMP-7. *Science* **265**, 1234–1237 (1994).
- Preckel, T. et al. Impaired immunoproteasome assembly and immune responses in PA28^{-/-} mice. *Science* **286**, 2162–2165 (1999).
- Anderson, K. et al. Endogenously synthesized peptide with an endoplasmic reticulum signal sequence sensitizes antigen processing mutant cells to class I-restricted cell-mediated lysis. *J. Exp. Med.* **174**, 489–492 (1991).
- Paz, P., Brouwenstijn, N., Perry, R. & Shastri, N. Discrete proteolytic intermediates in the MHC class I antigen processing pathway and MHC I-dependent peptide trimming in the ER. *Immunity* **11**, 241–251 (1999).
- Shastri, N. & Gonzalez, F. Endogenous generation and presentation of the OVA peptide/K^b complex to T-cells. *J. Immunol.* **150**, 2724–2736 (1993).
- Cascio, P., Hilton, C., Kisselev, A. F., Rock, K. L. & Goldberg, A. L. 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J.* **V20**, 2357–2366 (2001).
- Toes, R. E. M. et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.* **194**, 1–12 (2001).
- Egen, J. G. & Allison, J. P. Cytotoxic T lymphocyte antigen-4 accumulation in the immunological synapse is regulated by TCR signal strength. *Immunity* **16**, 23–35 (2002).

Acknowledgements

We thank J. Egen for help with fluorescence microscopy; D. King for peptide synthesis; and S. Schwab for comments. This work was supported by grants to N.S. from the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.S. (e-mail: nshastri@socrates.berkeley.edu).

Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex

Valentin I. Gordeliy^{*†}, Jörg Labahn^{*}, Rouslan Moukhametziev^{*†}, Rouslan Efremov^{*†}, Joachim Granzin^{*}, Ramona Schlesinger^{*}, Georg Büldt^{*}, Tudor Savopoli^{§||}, Axel J. Scheidig[§], Johann P. Klare[§] & Martin Engelhard[§]

^{*}Research Centre Jülich, Institute of Structural Biology (IBI-2), 52425 Jülich, Germany

[†]Centre for Biophysics and Physical Chemistry of Supramolecular Structures, MIPT, 141700, Moscow District, Russia

[§]Max-Planck-Institut für Molekulare Physiologie, Otto Hahn Str. 11, 44227 Dortmund, Germany

^{||}Present address: Carol Davila Medical and Pharmaceutical University, POB15-205 Bucharest, Romania.

Microbial rhodopsins, which constitute a family of seven-helix membrane proteins with retinal as a prosthetic group, are distributed throughout the Bacteria, Archaea and Eukaryota^{1–3}. This family of photoactive proteins uses a common structural design for two distinct functions: light-driven ion transport and phototaxis. The sensors activate a signal transduction chain similar to that of the two-component system of eubacterial chemotaxis⁴. The link between the photoreceptor and the following cytoplasmic signal cascade is formed by a transducer molecule that binds tightly and specifically⁵ to its cognate receptor by means of two transmembrane helices (TM1 and TM2). It is thought that light excitation of sensory rhodopsin II from *Natronobacterium pharaonis* (SRII) in complex with its transducer (HtrII) induces an outward movement of its helix F (ref. 6), which in turn triggers a rotation of TM2 (ref. 7). It is unclear how this TM2 transition is converted into a cellular signal. Here we present the X-ray structure of the complex between *N. pharaonis* SRII and the receptor-binding domain of HtrII at 1.94 Å resolution, which provides an atomic picture of the first signal transduction step. Our results provide evidence for a common mechanism for this process in phototaxis and chemotaxis.

Crystallization of the receptor–transducer complex, a member of the two-component signalling cascade (Fig. 1), has been achieved successfully using a shortened transducer (residues 1–114; *N. pharaonis* HtrII₁₁₄) comprising the two transmembrane helices (TM1 and TM2) and an additional small cytoplasmic fragment. This construct satisfies the properties of an appropriate model system for the native receptor–transducer complex as indicated by a low dissociation constant ($K_d \approx 100$ nM, S. Hippler-Mreyen, unpublished data) and by its capability to inhibit the inherent proton pump activity of SRII, as was shown for a larger transducer fragment (G. Schmies, unpublished data) and the full-length transducer^{8,9}. These data and those establishing a functional signal transfer from receptor to transducer⁷ indicate that HtrII₁₁₄ forms a functionally unimpaired complex with its cognate receptor SRII.

The thin orange crystals of SRII in complex with HtrII₁₁₄ grown in lipidic cubic phase¹⁰ displayed an orthorhombic shape of about 140 µm in size and diffracted to 1.8 Å. The asymmetric unit contains one complex. The expected dimer of the complex is formed by a crystallographic two-fold rotation axis, which is located in the middle of four transmembrane helices: TM1, TM2, TM1', TM2' (where a prime indicates the right-hand complex; Fig. 2a). The transmembrane helices F and G of the receptor are in contact with the helices of the transducer. The overall X-ray structure is in good agreement with a recently published model of the receptor–trans-

ducer complex deduced from electron paramagnetic resonance (EPR) measurements⁷.

The structure of SRII complexed with its transducer is markedly similar to that of the receptor alone including the retinal conformation^{11,12}. Obviously, the binding of the transducer to helices F and G hardly interferes with the side-chain arrangement of the receptor. A notable exception is found for Tyr 199. The aromatic plane of Tyr 199 has turned in the complex by about 90° and is now pointing into the direction of TM2 where its phenolic group forms a hydrogen bond to Nδ(2)-Asn 74 (2.8 Å). An interaction of Tyr 199 with the transducer has been proposed previously¹². It should be mentioned that a chloride ion, identified by ref. 11, close to the guanidinium group of Arg 72 is clearly absent in the present structure. The crystallization conditions used by this study (low pH and high NaCl concentration) favour the uptake of a chloride ion and therefore can explain the differences. The natural habitat of *N. pharaonis* (pH > 9) does not support the binding of a chloride ion to SRII.

The interface between receptor and transducer is formed mainly by van der Waals (vdW) contacts and only a few hydrogen bonds. Whereas the straight TM2 is oriented parallel to helix G of the receptor, TM1 is kinked at Gly 37 and bends away from the receptor. Thus, a crevice (formed by helices A, G, TM2 and TM1) opens to the cytoplasmic surface (Fig. 2a), which might harbour the back-folded amino terminus of TM1, as residual electron densities suggest. Although only van der Waals contacts are observed between the four transducer helices themselves, defined cross-connections are observed between receptor and transducer. The F–G loop region affixes the transducer by several contacts as well as by three hydrogen bonds between Thr 189 (SRII), Glu 43 (TM1) and Ser 62 (TM2) (Fig. 3). A second anchor point is observed in the middle of the membrane where, as mentioned above, the phenolic hydroxyl of Tyr 199 (helix G) bridges to Asn 74 (TM2). A view from the cytoplasm down the binding domain (Fig. 4a) reveals that closer contacts are between helix G and TM2, fixating these two transmembrane helices to one another. There are twice as many van der Waals contacts between helices G and TM2 than between F and TM2. The closer packing between G and TM2 can be quantified by an average van der Waals distance of 4.06 Å in comparison to a value of 4.22 Å between F and TM2 (Fig. 4b). The four helices of the

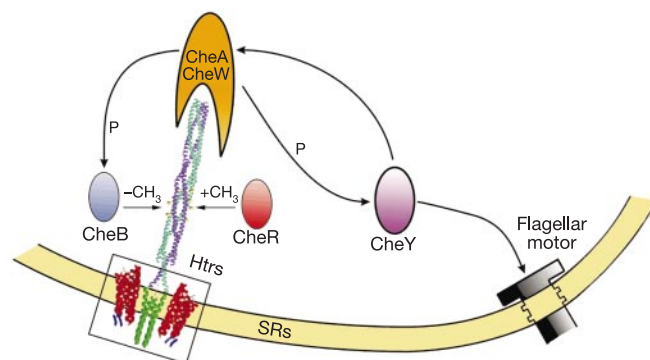


Figure 1 Two-component signalling cascade. The activation of the transducer HtrII by its receptor SRII leads to a conformational change of TM2 that propagates to the tip of the coiled-coil cytoplasmic domain (structure taken from ref. 31). The next steps in the signalling cascade involve—in analogy to the bacterial sensory system²⁰—the homodimeric histidine kinase CheA, the coupling protein CheW, and the response regulators/aspartate kinases CheY and CheB. Phosphorylated (P) CheY functions as a switch factor of the flagellar motor. CheB (a methyltransferase) together with CheR (a methyltransferase) are involved in the adaptation processes of the bacteria. The box highlights the receptor–transducer complex reported in the present study. SR, sensory rhodopsin.

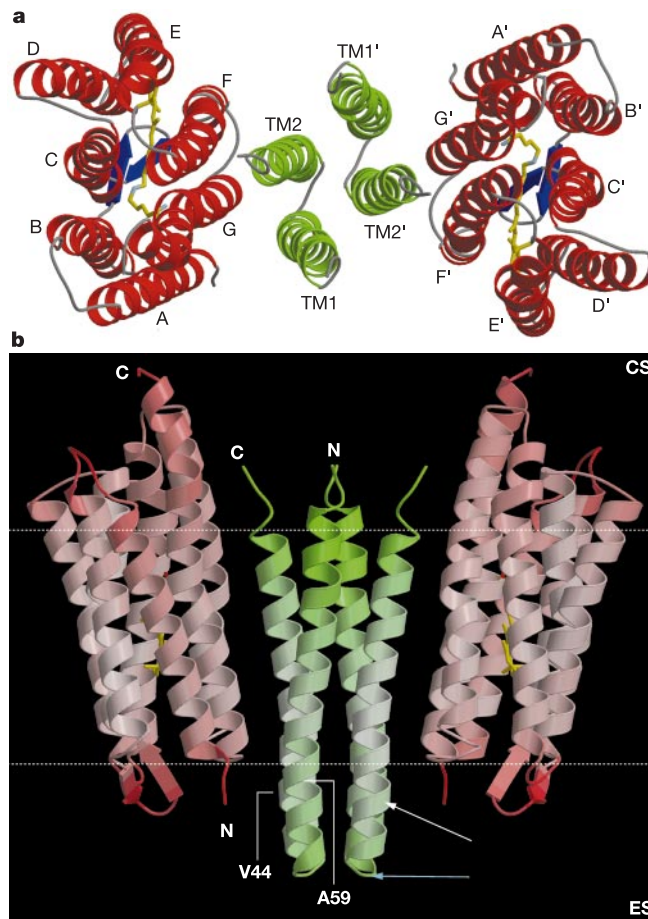


Figure 2 Fold of the receptor-transducer complex **a**, Ribbon diagram of the top view from the cytoplasmic side. α -Helices are in red for the receptor and green for the transducer; β -strands are in blue and coils in grey. The labels of the symmetry related complex are marked by a prime. The crystallographic symmetry axis is located between TM1–TM2 and TM1'–TM2'. **b**, Side view of the complex. The complex is coloured according to *B*-factor mobility: light red/green (less mobile), dark red/green (mobile). ES, extracellular side; CS,

cytoplasmic side. The dotted white lines confine the major hydrophobic core of the proteins. Of note, the actual membrane boundary will not follow these straight lines. The arrows indicate the shortened stalk in HtrI (white) and the site where the helices α 1 and α 4 of the chemoreceptor domain of *H. salinarum* HtrII are attached to the transmembrane helices TM1 and TM2 (blue). All figures were generated with MOLSCRIPT and Raster3D.

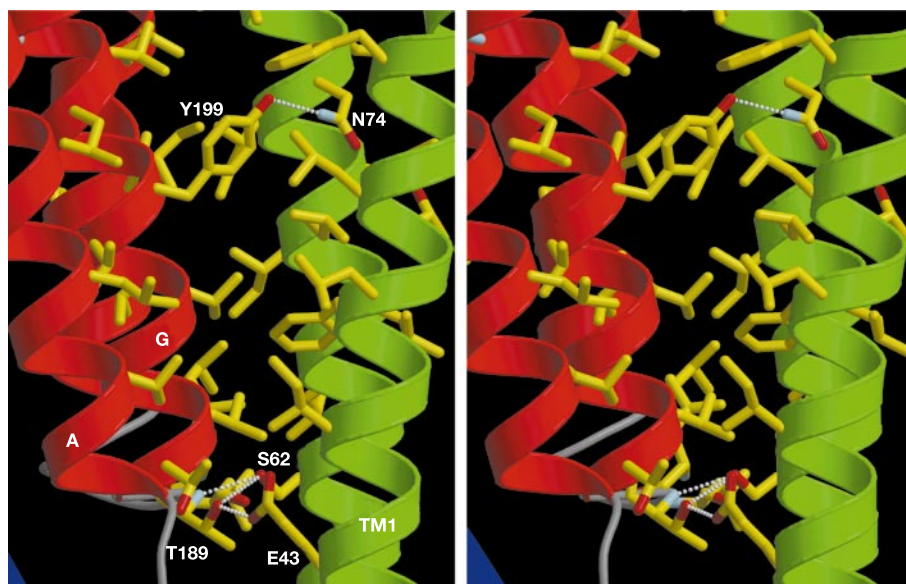


Figure 3 Stereo view of the hydrogen bonds and van der Waals contacts between receptor (α -helices in red) and transducer (α -helices in green). The residues that are involved in hydrogen bonds are labelled.

transducer in the dimer are packed against each other, thereby intercalating their bulky hydrophobic side chains.

Further insight into the structure of the receptor–transducer complex is gained if the structure is viewed perpendicular to the membrane. Most notably, the four helices of the transducer extend beyond the extracellular side by about three helical turns, comprising residues 44–59 (Fig. 2b). This sequence is missing in HtrI from *Halobacterium salinarum*, as obtained from a sequence alignment with *N. pharaonis* HtrII, but is repeated at the N-terminal end of TM2 in *H. salinarum* HtrII (ref. 13). Notably, *H. salinarum* HtrII not only transmits the signal from the photoreceptor SRII but also operates as a chemoreceptor¹⁴. This function is conferred by a serine-binding domain that is inserted in front of the sequence forming the stalk. Crystal structures of the ligand-binding domain of homologous eubacterial aspartate receptors display two extended helices in the dimer ($\alpha 4$ and $\alpha 4'$), which could connect to a structural element like the stalk¹⁵. The observation of different degrees of periplasmic domain excision is in line with the evolution of the four archetypical halobacterial rhodopsins, which has been explained by two gene duplication events¹⁶. According to this hypothesis a proto-chemoreceptor gene has been acquired by the proto-sensor gene after the first duplication step.

The temperature factors along the transducer and the receptor helices F and G increase from the centre to the cytoplasmic side of the proteins (Fig. 2b). Furthermore, the structure of TM2 ends with Leu 82, leaving residues 83–114 in multiple conformations, although they are necessary for receptor binding (S. Hippler-Mreyen, personal communication). An increased flexibility in this part of the molecule would facilitate the shift between the active and inactive conformations, which should be thermodynamically close to each other.

As shown previously, light excitation of *N. pharaonis* SRII triggers an outward movement of the cytoplasmic part of helix F (first observed for bacteriorhodopsin^{17,18}). It was proposed that this conformational change induces a clockwise rotation of TM2 if seen from the cytoplasmic side (ref. 7). On the other hand, the binding of ligands to the chemotaxis receptor domain generates a piston-type sliding of one of its helices ($\alpha 4$) of about 1.6 Å (although a small rotation can not be excluded)^{19,20}. It was proposed that this conformational change propagates along TM2—which is joined to $\alpha 4$ —towards the inside of the cell (reviewed by ref. 20). As

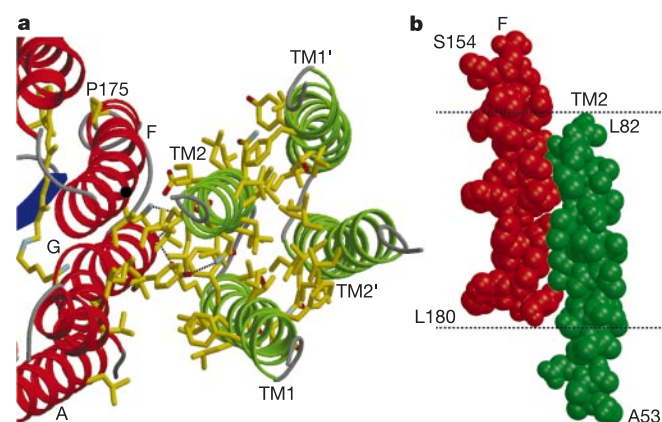


Figure 4 Interactions between SRII and HtrII. **a**, View from the cytoplasm down the binding domain. Several residues involved in the van der Waals contact are shown. The transducer helices are coloured green. The black dot on helix F marks the equivalent height in helix F corresponding to the start of helix TM2. **b**, Space-filling model showing close contact between helix F and TM2 as viewed from the direction of helix G. Starting and ending amino acids are shown. The black dashed lines confine the hydrophobic core as in Fig. 2b.

Table 1 X-ray crystallographic data

Data collection*	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions (Å)	124.30, 46.96, 53.84
Unique reflections; average redundancy	23,156; 3.5
Completeness (%)	94.7 (77.3)
R _{merge} (%)	6.2 (39.4)
I/σ(I)	5.5 (1.7)
Refinement†	
R _{work} ; R _{free} (%)	22.8; 25.8
Average B-factor (Å ²); No. of atoms	32.2; 2,187
r.m.s. deviation: bonds (Å); angles (°)	0.009; 1.07

$R_{\text{merge}} = \sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)$, where $I_i(h)$ is the i th measurement and $\langle I(h) \rangle$ is the mean of all measurements of $I(h)$. Numbers in parentheses indicate the value for the outermost shell. For calculation of R_{free} , 5% of the reflections were reserved.

$R = \sum_h ||F_o(h)| - |F_c(h)|| / \sum_h |F_o(h)|$; where $|F_o|$, $|F_c|$ are observed and calculated structure factor amplitudes.

*Data were collected at beamline ID14-1 of ESRF at wavelength 0.934 Å.

†Residue range: SRII, 1–225; Htr, 24–82.

the *H. salinarum* SRII–HtrII complex displays dual functionality, a common mechanism of transmembrane signalling for phototaxis and chemotaxis should exist. The crystal structure of *N. pharaonis* SRII–HtrII reveals features that bear significance for the process of signal propagation across the membrane.

The present structure and the observation that the flap-like movement of helix F induces a rotation of the cytoplasmic end of TM2 (refs 6, 7), gives rise for a probable mechanism of transmembrane signalling. Helix F contains a proline residue (Pro 175; Fig. 4a) at an equivalent depth as Tyr 199 (Figs 3 and 4a), which could function as a hinge for the light-activated movement of helix F (a similar role has been proposed for Pro 185 in bacteriorhodopsin²¹). If the outward bending of helix F is in the same direction as observed for bacteriorhodopsin, it should collide tangentially with TM2 (Figs 2a and 4), thereby inducing its rotary motion. A primary piston-like transition of TM2 seems unlikely because the structure shows that TM2 has numerous side-chain intercalations with helix G and TM1 (see above and Fig. 4a). As TM2 is anchored to helix G, the rotation might occur around an axis located between the two helices. Rotation might involve an unwinding of the cytoplasmic coiled-coil helices (Fig. 1) as was observed for the γ -subunit of the F_1F_0 -ATP synthase rotary motor²². As mentioned above, signal transfer from chemoreceptors and photoreceptors should follow similar mechanisms. A screw-like movement would satisfy both observations of a piston as well as a rotary motion of TM2. It should be noted that the initial ligand-induced shift of TM2 (whether a piston, rotary, or screw-type motion) might just be the trigger to shift the equilibrium from the inactive to the active state of the transducer, thereby producing a similar final conformation, recognized by the His kinase CheA.

Another interesting aspect derived from the structure of the receptor–transducer complex concerns the negative cooperativity observed not only for chemoreceptors but also, for example, for tyrosine kinase receptors such as the insulin receptor (discussed by ref. 23). Before excitation, the environment of the two helices TM2 and TM2' appears to be the same (Fig. 2a). After excitation of one receptor molecule the corresponding TM2 will move, thus changing the environment of the other TM2, which might inhibit its equivalent movement. Therefore, the photoreceptors would show the negative cooperativity characteristic for chemoreceptors.

The high-resolution structure of the receptor–transducer complex provides the foundation for understanding transmembrane signalling on a molecular level. It has now become possible to solve the structure of intermediates, as has been done for bacteriorhodopsin. This new information will be of significance not only for bacterial phototaxis and chemotaxis but also for other dimeric receptors, which might lead finally to a general model of transmembrane signal transduction. □

Methods

Protein preparation

The genes of *N. pharaonis* SRII and the carboxy terminal truncated transducer (1–114) were cloned into a pET27bmod expression vector²⁴ with a C-terminal $\times 7$ His tag, respectively. Proteins were expressed in *Escherichia coli* strain BL21 (DE3), and purified as described^{25,26}. After removal of imidazol by diethyl-aminoethyl chromatography, SRII-His and HtrII₁₁₄-His were mixed in a 1:1 ratio, followed by reconstitution into purple membrane (the bacteriorhodopsin containing membrane patches of *H. salinarum*) lipids⁷ (protein to lipid ratio 1:35). After filtration, the reconstituted proteins were pelleted by centrifugation at 100,000g. For resolubilization, the samples were resuspended in a buffer containing 2% *n*-octyl- β -D-glucopyranoside and shaken for 16 h at 4 °C in the dark. The resolubilized complex was isolated by centrifugation at 100,000g.

Crystallization, structure determination and refinement

We added the solubilized complex in crystallization buffer (150 mM NaCl, 25 mM Na/KPi, pH 5.1, 0.8% *n*-octyl- β -D-glucopyranoside) to the lipidic phase, formed from monovaccenin (Nu-Chek Prep). Precipitant was 1 M salt Na/KPi, pH 5.6. Crystals were grown at 22 °C.

X-ray diffraction data were collected at beamline ID14-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, using a Quantum ADSC Q4R CCD (charge-coupled device) detector. Data were integrated using MOSFLM²⁷ and SCALA²⁸. Molecular replacement using MOLREP²⁸ to phase a polyaniline model (from Protein Data Bank accession number 1JGJ (ref. 12)) gave a unique solution ($R = 0.568$, correlation coefficient $C = 0.357$) at 2.9 Å. After inserting side chains for SRII, the helices of HtrII were found ($R = 0.329$, $C = 0.711$). Simulated annealing, positional refinement and temperature factor refinement were performed in CNS²⁹; model rebuilding was carried out in O³⁰ (Table 1).

Received 9 May; accepted 16 July 2002; doi:10.1038/nature01109.

- Oesterhelt, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature* **233**, 149–152 (1971).
- Bieske, J. A. *et al.* A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. *Biochemistry* **38**, 14138–14145 (1999).
- Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Rudolph, J. & Oesterhelt, D. Deletion analysis of the che operon in the archaeon *Halobacterium salinarum*. *J. Mol. Biol.* **258**, 548–554 (1996).
- Zhang, X. N., Zhu, J. & Spudich, J. L. The specificity of interaction of archaeal transducers with their cognate sensory rhodopsins is determined by their transmembrane helices. *Proc. Natl Acad. Sci. USA* **96**, 857–862 (1999).
- Wegener, A. A., Chizhov, I., Engelhard, M. & Steinhoff, H. J. Time-resolved detection of transient movement of helix F in spin-labelled pharaonis sensory rhodopsin II. *J. Mol. Biol.* **301**, 881–891 (2000).
- Wegener, A. A., Klare, J. P., Engelhard, M. & Steinhoff, H. J. Structural insights into the early steps of receptor-transducer signal transfer in archaeal phototaxis. *EMBO J.* **20**, 5312–5319 (2001).
- Schmies, G. *et al.* Electrophysiological characterization of specific interactions between bacterial sensory rhodopsins and their transducers. *Proc. Natl Acad. Sci. USA* **98**, 1555–1559 (2001).
- Sasaki, J. & Spudich, J. L. Proton circulation during the photocycle of sensory rhodopsin II. *Biophys. J.* **77**, 2145–2152 (1999).
- Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases—a novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
- Royant, A. *et al.* X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proc. Natl Acad. Sci. USA* **98**, 10131–10136 (2001).
- Luecke, H. *et al.* Crystal structure of sensory rhodopsin II at 2.4 Å: insights into colour tuning and transducer interaction. *Science* **293**, 1499–1503 (2001).
- Zhang, W. S., Brooun, A., Mueller, M. M. & Alam, M. The primary structures of the archaeon *halobacterium salinarum* blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein. *Proc. Natl Acad. Sci. USA* **93**, 8230–8235 (1996).
- Hou, S. B. *et al.* Sensory rhodopsin II transducer HtrII is also responsible for serine chemotaxis in the archaeon *halobacterium salinarum*. *J. Bacteriol.* **180**, 1600–1602 (1998).
- Yeh, J. I. *et al.* High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. *J. Mol. Biol.* **262**, 186–201 (1996).
- Ihara, K. *et al.* Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. *J. Mol. Biol.* **285**, 163–174 (1999).
- Koch, M. H. J. *et al.* Time-resolved X-ray diffraction study of structural changes associated with the photocycle of bacteriorhodopsin. *EMBO J.* **10**, 521–526 (1991).
- Subramaniam, S., Gerstein, M., Oesterhelt, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.* **12**, 1–8 (1993).
- Lynch, B. A. & Koshland, D. E. Jr Structural similarities between the aspartate receptor of bacterial chemotaxis and the trp repressor of *E. coli*. Implications for transmembrane signaling. *FEBS Lett.* **307**, 3–9 (1992).
- Falke, J. J. & Hazelbauer, G. L. Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem. Sci.* **26**, 257–265 (2001).
- Gerwert, K., Hess, B. & Engelhard, M. Proline residues undergo structural changes during proton pumping in bacteriorhodopsin. *FEBS Lett.* **261**, 449–454 (1990).
- Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F₁-ATPase

from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).

- Koshland, D. E. Jr The structural basis of negative cooperativity: receptors and enzymes. *Curr. Opin. Struct. Biol.* **6**, 757–761 (1996).
- Klostermeier, D., Seidel, R. & Reinstein, J. Functional properties of the molecular chaperone DnaK from *Thermus thermophilus*. *J. Mol. Biol.* **279**, 841–853 (1998).
- Hohenfeld, I. P., Wegener, A. A. & Engelhard, M. Purification of histidine tagged bacteriorhodopsin, pharaonis halorhodopsin and pharaonis sensory rhodopsin II functionally expressed in *Escherichia coli*. *FEBS Lett.* **442**, 198–202 (1999).
- Shimono, K., Iwamoto, M., Sumi, M. & Kamo, N. Functional expression of pharaonis phorbodhodopsin in *Escherichia coli*. *FEBS Lett.* **420**, 54–56 (1997).
- Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image data. *CCP4 ESF-EACMB Newslett. Protein Crystallogr.* **26** (2002).
- Collaborative Computational Project The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kim, K. K., Yokota, H. & Kim, S. H. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400**, 787–792 (1999).

Acknowledgements

We thank C. Baeken, I. Ritter and M. Schumacher for technical help, and B. Gehrman for secretarial assistance. Discussions with R. G. Goody are gratefully acknowledged. We also thank the support by the staff of beamline ID14-1, and in particular E. Mitchell, ESRF, Grenoble, France; H. J. Brandt and C. Wandrey (IBT-Jülich) for a high yield fermentation of SRII containing *E. coli* cells; A. K. Islamov, A. I. Kuklin and G.N. Bobarikina for help in investigations on mechanisms of membrane protein crystallization in lipidic phases; and I. N. Groznov and V. B. Kireev for their support of this work. This study was supported by the Deutsche Forschungsgemeinschaft, the Max-Planck-Gesellschaft, and the Alexander von Humboldt Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.B. (e-mail: g.bueltdt@fz-juelich.de) or M.E. (e-mail: martin.engelhard@mpi-dortmund.mpg.de). Coordinates have been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank under accession code 1H25.

correction

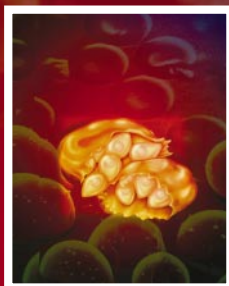
Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*

Jianbing Mu, Junhui Duan, Kateryna D. Makova, Deirdre A. Joy, Chuong Q. Huynh, Orale H. Branch, Wen-Hsiung Li & Xin-zhuan Su

Nature **418**, 323–326 (2002).

It has been drawn to our attention that the 3D7 parasite, labelled as an African parasite in our Letter, was initially isolated from The Netherlands, although we have some genetic evidence suggesting an African origin (Wootton, J.C. and Su, X., manuscript in preparation). Therefore, 3D7 in our paper should be labelled as having a European origin, with a good probability of having originated from Africa. □

Good 'omics' for the poor?



Cover illustration
Infected red blood cells liberating *P. falciparum* merozoite parasites. (Image credit: Debbie Maizels.)

This week marks a milestone in malaria research, with the publication of complete genome sequences for the human parasite, the apicomplexan *Plasmodium falciparum* (this issue of *Nature*), and its vector, the mosquito *Anopheles gambiae* (*Science*, 4 October). Some may warn against excessive optimism, because the global problems caused by malaria are daunting (see Malaria Insight, *Nature* 415, 669–715, 2002; News and Views, this issue, pp. 493–497; News Features, pp. 426–430). As an antidote to pessimism, let us celebrate the science you will find in this section, which will without doubt aid researchers in the fight against malaria.

Plasmodium falciparum is the first eukaryotic parasite for which we have a complete genome (pp. 498, 527, 531 & 534). The large proportion (70%) of predicted genes already validated experimentally allow firm conclusions to be drawn about the evolution of metabolic pathways. Comparison with the human genome also reveals some pathways that are specific to the pathogen or its peculiar organelles; these will usher in the development of specific drugs with lesser side effects. Completion of a second full genome, that of the model rodent malaria parasite *P. yoelii yoelii*, allows, for the first time, the comparison of two eukaryotic species within a single genus: *Plasmodium* (page 512). Despite their evolutionary similarity, these pathogens exhibit striking differences in their immune evasion strategies. The immediate availability of two state-of-the-art proteomics studies provides stimulating new insights for the development of both drugs and vaccines (pp. 520, 537). Newly discovered patterns of gene expression during the *Plasmodium* life cycle will lead to strategies for targeting several parasitic stages at once.

The fruits of applying this knowledge may take years to materialize, so could this be just another end of a beginning? We believe not. This major achievement will maintain the momentum in the scientific community worldwide. Researchers have already used the freely available PlasmoDB database (p. 490) to identify new potential antimalarial drugs (*Nature Medicine* 7, 167; 2001). In the same spirit, all of this section's contents, along with seminal malaria research, news and features articles previously published in *Nature*, are available free online (• www.nature.com/nature/malaria). A CD-ROM containing similar items, plus an interactive GenePlot from PlasmoDB, will be distributed with a future issue of *Nature*, bringing this wealth of information to researchers in countries with limited Internet access. That high-tech genomics and proteomics are being mobilized against the emblem disease of poverty is a good omen indeed.

Tanguy Chouard Senior Editor
Ursula Weiss Senior Editor
Ritu Dhand Chief Biology Editor

commentary

- 490 The *Plasmodium* genome database

news and views

- 493 The parasite genome: The grand assault
- 495 The parasite genome: Biological revelations

- 496 The mosquito genome: The post-genomic era opens

articles

- 498 Genome sequence of the human malaria parasite *Plasmodium falciparum*
- 512 Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*
- 520 A proteomic view of the *Plasmodium falciparum* life cycle

letters to nature

- 527 Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13
- 531 Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14
- 534 Sequence of *Plasmodium falciparum* chromosome 12
- 537 Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry

The *Plasmodium* genome database

Designing and mining a eukaryotic genomics resource.

The *Plasmodium* Genome Database Collaborative

As reported elsewhere in this issue (M. J. Gardner *et al.* *Nature* **419**, 498–511; 2002), a reference genome sequence for the human malaria parasite *Plasmodium falciparum* is now complete. But how are researchers to access *P. falciparum* genome sequence data, integrate this resource with other relevant data sets, and exploit the resulting information for functional studies, including identification of novel drug targets and candidate vaccine antigens?

The *Plasmodium* genome database (PlasmoDB, see <http://PlasmoDB.org>) contains information from multiple sources, including DNA sequence data and curated annotations, automated gene model predictions, predicted proteins and protein motifs, cross-species comparisons, optical and genetic mapping data, information on population polymorphisms, expression data generated by a variety of complementary strategies, and proteomics data. Integrating this information at a single site provides 'one-stop shopping' for genomics-scale data sets related to malaria parasites.

The use of a relational database architecture enables users to ask complex questions. For example, immunologists trying to develop an antimalaria vaccine might wish to identify potential immunodominant surface antigens. Drug developers might wish to identify enzymes expressed in bloodstream parasites that differ significantly from their human counterparts. Researchers interested in antigenic variation and how the parasite adheres to cells (a cause of malaria pathogenesis) might wish to identify all gene families in the parasite genome; those interested in genome organization might be interested in the chromosomal location of these proteins; evolutionary biologists might wish to examine all genes for which clear orthologues are known from a range of species; and so on.

Universal access

It has taken six years to complete the *P. falciparum* genome sequence. In the meantime, interim data were periodically released by the three sequencing centres involved in this project, to advance research on basic malaria biology, and drug and vaccine development. PlasmoDB was developed to make this information available to the research community, notwithstanding the challenges posed by unfinished sequence data. This web-accessible database provides access to the entire genome sequence of the 3D7 reference strain of *P. falciparum*, together with computationally predicted and manually curated genes and gene models, protein feature predictions and functional annotation.

PlasmoDB went live in June 2000 — more than two years before today's formal completion of the *P. falciparum* reference sequence. The website receives several thousand hits each day from more than 100 countries, numbers that are certain to rise significantly with the release of the complete genome sequence. The result can be measured in the scores, possibly hundreds, of publications that have resulted, and in new

targets now being assessed for drug and vaccine development.

Malaria biologists are a more diverse and dispersed community than those who study fruitfly or yeast genomes. They encompass field scientists in Cameroon, epidemiologists in Papua New Guinea, pharmaceutical developers in India, molecular geneticists in Brazil, and so on. Because many malaria researchers lack reliable high-speed Internet access, a platform-independent CD-ROM (to be distributed with *Nature* in a few weeks' time) has been developed to provide universal free access to the complete genome sequence and annotations currently available for this malaria parasite. More than a series of 'flat-file' images, *P. falciparum* GenePlot is a true database, providing a graphical user interface for browsing, querying, downloading and manipulating the genome and annotations on a desktop computer without web access.

It has been a stimulating challenge to see how many commonly asked questions can be accommodated in the CD-ROM format. For example, while local implementation of BLAST searches requires substantial memory and computational speed (and GenBank is too large to include on a single CD), GenePlot can be asked to find and retrieve all predicted proteins with similarity to proteases, based on text indices derived from precomputed BLAST comparisons of the entire *P. falciparum* genome against all of GenBank.

The initial motivation behind the GenePlot CD was to make the genome accessible to malaria biologists with limited Internet connectivity, but this format has also proved enormously popular with well-connected users. Having the data literally 'in hand' provides scientists everywhere with a sense of ownership and involvement in the *Plasmodium* genome project, expediting the pace of research and discovery related to malaria parasites and the devastating diseases they cause.

Unfinished business

In most genomics projects, initial mapping studies (desirable even with the advent of whole-genome shotgun sequencing) are followed by a random sequencing phase, then by a phase focusing on closure of remaining gaps to produce a 'finished' sequence (which may still contain numerous gaps, depending on complexity and size of the genome, time, patience and funding). Annotation is conducted to various levels of depth. Database development makes the information accessible to the user community. Finally, functional studies (transcript profiling, proteomic studies, genome-scale knockouts, and so on) become possible once the complete, annotated sequence is available to end-users.

There are good reasons for this sequential strategy. Gap closure is expensive, and so makes little sense while random sequencing may still yield useful information. Manual annotation of assembled sequences is also laborious, and is best deferred until the genome sequence is complete. For large, complex eukaryotic genomes, years may pass between the initial sequencing and the availability of this information in practical form for researchers in the lab. Such delays cause considerable frustration, as individual genes could be identified long before assembly of a finished genome.

Problems associated with unfinished data, and the accompanying need for user education regarding the interpretation of these results, provided the first challenge for PlasmoDB. Specific information missing in incomplete data

The CD-ROM containing *P. falciparum* GenePlot and other malaria-related resources, including *Nature's* malaria Insight of 7 February 2002 and the papers reported elsewhere in this issue, will be provided to all *Nature* subscribers in a few weeks' time. It can also be obtained from helpcd@plasmodb.org or the Malaria Research and Reference Reagent Resource Center (MR4) by an e-mail request to malaria@atcc.org, with "Nature malaria CD-ROM" in the subject line. A full postal address must be included in the body of the message.

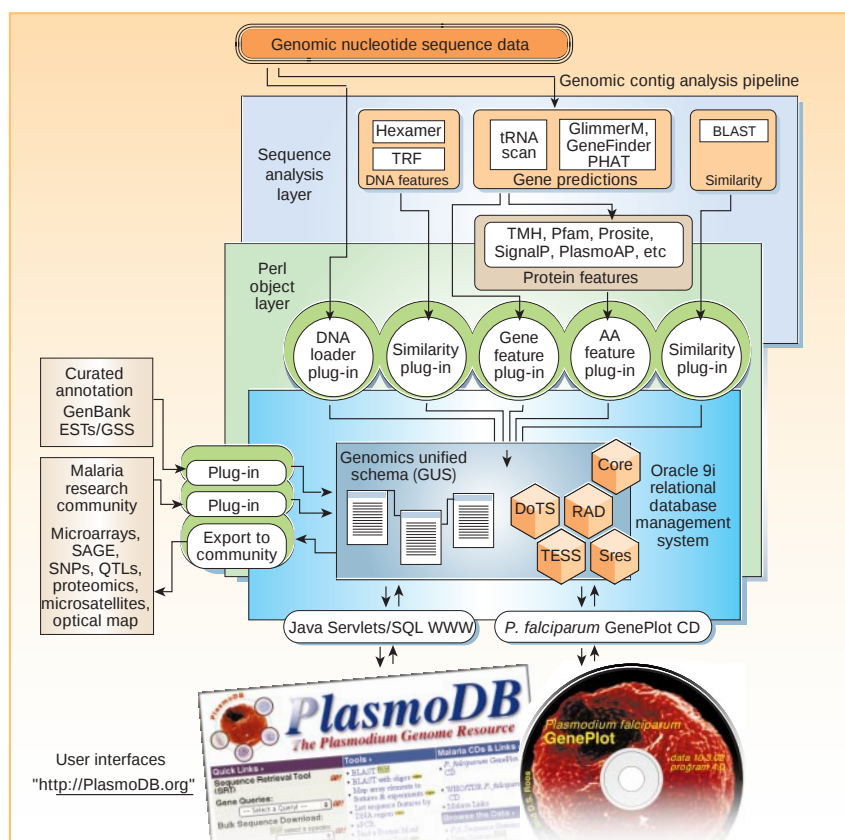
sets limits confidence that a particular gene is absent from the organism. Contaminating sequences from cloning vectors and host cells may be present. Redundancy in the data set attributable to incomplete or inaccurate assemblies poses a further problem, particularly for the A/T-rich *P. falciparum* genome. In PlasmoDB, possible redundancy or inaccurate assembly was identified by high-stringency comparisons of each sequence with the entire genome; and comparison of DNA sequences with optical and genetic maps. The importance of these tools for *P. falciparum* declines as the genome project approaches completion, but they remain valuable for new projects, such as the other *Plasmodium* species now being sequenced.

Unfinished sequence data also pose challenges for gene identification and analysis, as the constantly changing nature of this information makes time-consuming manual annotation impossible. Comparisons with GenBank, computational gene-finding algorithms and protein feature analyses are feasible (Box 1), but generate a bewildering range of predictions: which of four competing gene predictions is most likely to be correct? Which of 60 sequences exhibiting similarity to cathepsin is really a protease? Automated analysis can help to provide provisional assignments early, before manual curation of the finished sequence. Even after first-pass annotation, these analyses can help to suggest alternative possibilities whenever new experimental information suggests inaccuracies in the curated annotation.

Integrative 'omics'

Many disciplines accommodate large data sets (MRI imaging, weather forecasting, ecological and econometric modelling, and so on), but this is a relatively new problem for molecular and cell biologists. How to collect the deluge of data engulfing us from genomics, transcriptomics, proteomics, glycomics, pharmacogenomics, vaccinomics, and even more hideously named approaches? What kind of tools will be required to analyse — and to integrate — these massive 'omics'-scale data sets? How can we use all this information to treat malaria?

PlasmoDB is based on a relational database architecture (GUS; Box 1), built around biologically relevant relationships following the central dogma of biology: 'gene to messenger RNA to protein'. Parallel views for other organisms (including other *Plasmodium* species) allow phylogenetic comparisons. Because all this information is in a single database, queries can combine searches for particular genes of interest with RNA and protein expression analysis, studies on population genetic polymorphisms, and cross-species comparison. One can envisage the incorporation of other data types, such as publication records, clinical outcome data, genomic information from the mosquito vector *Anopheles gambiae*, protein structural information



Box 1: The architecture of PlasmoDB

PlasmoDB is not itself a database, but a web interface that uses an underlying relational database (GUS, for genomics unified schema), which stores and integrates nucleotide sequences, annotation, information on gene expression and regulation, controlled vocabularies/ontologies, and evidence for these annotations. GUS is organism-independent and also contains the human and mouse genomes (www.allgenes.org). The schema, associated code and project-independent data are at www.gusdb.org.

Primary *P. falciparum* sequence data are subjected to automated analyses (sequence analysis layer), including the identification of motifs and simple repeats; comparison against the entire genome to identify gene families, repetitive elements and redundancy; searching for intron/exon structure, using several algorithms trained on experimentally validated *P. falciparum* sequences; conceptual gene translation and identification of potential protein motifs; and

comparisons with the non-redundant GenBank/EMBL database (results retained in a text-queryable index). Genomic contig sequences are aligned to optical restriction maps and microsatellite linkage groups using hidden Markov models for fragment length and ePCR.

The GUS schema employs views that are used in an object layer for parent-child relationships. To facilitate data loading, Perl was used to create a 'thin' object layer in which each relational table is treated as an object. GUS is partitioned into distinct name spaces. Core contains workflow tables, tracking how each row in the database is populated (data provenance). Sres (shared resources) contains controlled vocabularies and ontologies, such as taxonomy, anatomy and disease tables. TESS captures descriptions (grammar representations) for genetic regulatory regions (not currently implemented for PlasmoDB). DoTS houses sequence and sequence annotation. Any sequence span

can have multiple features mapped to it, and gene predictions can be associated with multiple transcripts and proteins. Each predicted or experimentally determined transcript may itself have multiple features and similarities, as can each protein entry. RAD handles data from high-throughput technologies for studying gene expression. RAD currently accommodates expression data from SAGE (serial analysis of gene expression), cDNA and oligonucleotide glass slide microarrays, and Affymetrix chips, and is extensible to accommodate information from other platforms. Sample information, together with other experimental descriptions, can be entered directly into the database via web-based forms.

The RAD schema is compliant with MIAME guidelines (www.mged.org/). A microarray gene expression (MAGE) object model and XML-based language have been developed for data exchange, and importers and exporters are being built for RAD to MAGE-ML.

from high-throughput crystallography studies, and chemical compound libraries.

The power of database queries

PlasmoDB provides graphic and text-based views of all available *Plasmodium* genomic sequences, curated annotations, and tools for retrieval of these data. But the sheer wealth of information can make browsing difficult, so the database allows the user to define custom views. For all their visual appeal, however, static, precomputed views are inherently restricted, and so fail to answer many genomic-scale questions that arise in the laboratory.

The relational database underlying PlasmoDB permits queries that integrate diverse data types, as illustrated by questions relating to drug and vaccine development (Table 1). For example, a medicinal chemist might be interested in *P. falciparum* dihydrofolate reductase (DHFR), the target of the drug pyrimethamine used in common antimalarial agents. The gene encoding this enzyme can be identified by EC number or GO function, text searches of curated annotation using the enzyme name as a key word, text searches against BLAST results, motif searches for protein sequence signatures, BLAST similarity to DHFR sequences from other species, or searches based on protein structural predictions. Degenerate searches are also possible, such as searching for all proteases. The results returned would undoubtedly contain false positives, but these can be weeded out by scientists familiar with protease characteristics. Candidate cytoskeletal proteins can be identified by similar strategies, or searches based on protein structural predictions. Such searches can then be refined, for example by identifying sequences conserved in multiple malaria parasites, or those that are sufficiently distinct from human orthologues to provide a basis for selective inhibition.

Information on metabolic pathways and/or subcellular localization can also be used to inform database queries. For example, PlasmoDB enables the identification of proteins likely to be associated with the apicoplast — a distinctive organelle that has received considerable attention as a candidate drug target — on the basis of curated annotation, exploiting the structured gene ontology (GO) vocabulary. Alternatively, the origins of this organelle by horizontal transfer of an algal chloroplast can be exploited as the basis for a text search for genes exhibiting sequence similarity to plastid, chloroplast or plant genes. Phylogenetic comparison with plant species is not currently supported in PlasmoDB, but all nucleotide and predicted protein sequences can be downloaded by users for local analysis.

Combining gene and protein predictions with the results from RNA and/or protein expression analysis enables enzymes being considered for antimalarial drug development to be filtered, removing any proteins not expressed in blood-stage parasites. Integrating

Table 1 Querying *Plasmodium* genome data

User query	Computational strategy/approach
Drug development	
Dihydrofolate reductase	GO function*, EC*, text*, motif* and BLAST* searches
Proteases	Text* and motif* searches; GO function or process*
Cytoskeletal genes conserved in multiple <i>Plasmodium</i> species	GO cellular component*; protein structural predictions; phylogenetic cross-comparison with other <i>Plasmodium</i> species*
Differ significantly from probable human orthologues	Comparison with human sequences*
Apicoplast pathway enzymes	GO function, process or component*; text search for 'chloroplast or plastid*'; phylogenetic comparison with plants/algae
Expressed in blood-stage parasites	Expression profiling studies*; proteomics data*
Essential for parasite survival	Curated annotation from pharmacological/genetic studies; literature searches
Validated drug targets (in other systems)	Drug databases; literature databases
Availability of candidate inhibitors	Small-molecule databases; DOCKing algorithms
Vaccine candidates	
Known antigens (AMA1, MSP1, SERA)	Text*, motif* and BLAST* searches
Multigene families	Self-BLAST analysis*
Associated with the parasite on infected cell surface	Protein features: signal sequences*, transmembrane domains*, potential acylation sites or GPI anchors. Similarity to known membrane and surface proteins in other systems
Immunodominant	B- and/or T-cell epitope predictions*
Unlikely to be deleted	Non-telomeric*; conserved in multiple <i>P. falciparum</i> isolates
Likely to be under positive (immune) selection	DNA/protein features: repetitive/low-complexity sequence*. High ratio of non-silent/silent polymorphisms (from phylogenetic cross-comparisons and population genetics studies)*

*Searches currently supported by PlasmoDB

these data with functional studies, polymorphism data, publications, or small-molecule databases, would allow further refinement.

For immunologists, computationally accessible queries allow identification of particular genes of interest as vaccine antigens (see Table 1). Additional gene-family members can be recognized on the basis of sequence similarity. Probable surface antigens can be identified from the presence of signal sequences, transmembrane domains, acylation signals or glycosylphosphatidylinositol (GPI) anchor motifs. Additional queries of immunological relevance might include the presence of predicted immunodominant epitopes, expression in life-cycle stage(s) of interest, conservation in multiple *P. falciparum* isolates, and evidence of immune selection based on highly repetitive elements, low-complexity sequence or polymorphisms identified in population genetic studies.

PlasmoDB can be used to build complex queries using boolean operators. For example, searching PlasmoDB release 3.3 for all genes predicted to contain a secretory signal sequence yields 1,952 hits. Because this search used curated annotations plus the predictions from any one of several distinct gene-finding algorithms, the results are several-fold redundant, yielding about 800 distinct genes, or more than 15% of the parasite genome. More than twice as many proteins (5,003) are predicted to contain transmembrane domains, but the intersection of these results yields only 1,083 hits (about 400 distinct proteins) exhibiting both features. Next, the database can be searched for all messenger RNAs known from expressed sequence tag (EST) evidence, yielding 3,057 hits (searches based on microarray or proteomics evidence are

also possible). The intersection between these secretory pathway and expression searches identifies a grand total of 190 candidates, probably corresponding to fewer than 100 distinct genes.

Two key points emerge from these queries. First, the power of a database devoted to mining genomics-scale data sets comes from its ability to form relational (integrated) queries, allowing researchers to frame their own questions. No encyclopaedic version of precomputed analyses and 'canned' queries will ever provide all possible answers in advance. For example, neither computational analysis nor manual curation would have been likely to identify enzymes associated with the apicoplast before this organelle was discovered and its targeting signals mapped.

Second, the goal of these queries is not to get the 'right' answer (a provably correct list of valid drug targets or vaccine antigens), but to reduce the options, filtering the overwhelming number of sequences in the genome down to a few genes amenable to experimental analysis — in short, to let computers do what computers do well, and to let people do what people do well. Integrating the results of such studies into the database completes the loop, with computational and experimental analysis in the lab building on each other to accelerate the pace of biological research. ■

Jessica C. Kissinger, Brian P. Brunk, Jonathan Crabtree, Martin J. Fraunholz, Bindu Gajria, Arthur J. Milgram, David S. Pearson, Jonathan Schug, Amit Bahl, Sharon J. Diskin, Hagai Ginsburg, Gregory R. Grant, Dinesh Gupta, Philip Labo, Li Li, Matthew D. Mailman, Shannon K. McWeeney, Patricia Whetzel, Christian J. Stoeckert Jr and David S. Roos are associated with the Departments of Biology and Genetics, Center for Bioinformatics and Genomics Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018, USA. Address for correspondence: droos@sas.upenn.edu

The grand assault

Russell F. Doolittle

The complete genome sequence of the parasite responsible for most of the world's human malaria has been determined. The nature of the genome meant that this was a difficult project, requiring considerable ingenuity.

The parasite *Plasmodium falciparum*, responsible for most human malaria, is among the most studied pathogens of all time, probably surpassed only by the human immunodeficiency virus and the tuberculosis bacterium *Mycobacterium tuberculosis*. The extent of human suffering caused by malaria and its devastating costs have long been recognized by international bodies, and many initiatives have been taken over the years to try to defeat this insidious microbe¹. In 1996, an international consortium of scientists from more than a dozen institutions set out to determine the 23 million base pairs of DNA that make up the organism's genome sequence. Their massive effort — which ended up going well beyond simple sequencing — is reported on pages 498–542 of this issue^{2–8}. The avowed goal of the project was to search for chinks in the parasite's armour, so that new and effective drugs and vaccines might be developed.

Sequencing strategy

The strategy for determining the *P. falciparum* genome sequence depended on first physically separating its 14 chromosomes by the technique of pulsed gel electrophoresis. In fact, three of the chromosomes (numbers 6, 7 and 8) could not be separated from each other and were simply taken as a combined unit. Three different teams then attacked different chromosomes: a team led by the Sanger Centre, Cambridge, UK, sequenced nine³; The Institute for Genomic Research (TIGR), Maryland, and others took on four⁴; and a group centred on Stanford University, California, did the other⁵.

In broadest outline, the DNA was mechanically sheared into random fragments, the fragments were inserted into bacteria (where they were copied every time the bacteria multiplied), and individual bacterial colonies were collected. The DNA inserts from these clones were sequenced automatically, and their order determined by assembling overlapping sequences together using a computer. Around half a million individual fragments were sequenced. As in most genome projects on this scale, the sequence determination is not really complete, and several gaps and ambiguities remain. Nonetheless, even at 95% 'finished', the published sequence must be regarded as a milestone that will be a major asset to biomedical researchers.

But why did this project take so long? After all, 23 million base pairs is not a particularly large genome by current standards (Fig. 1, overleaf), and the much larger genome of the fruitfly *Drosophila melanogaster* was apparently completed in less than a year⁹. The single biggest hurdle was the extremely biased base composition of the *P. falciparum* genome. More than 80% of the bases are either As or Ts (as opposed to Cs or Gs). In fact, regions of the genome that do not code for genes average more than 85% As or Ts, and runs of 50 As or Ts are common. Most of the genomes

already sequenced have been much less skewed in their base composition.

The extreme bias made the assembly process — by which individual clones are put in their correct order by an iterative overlap process — particularly challenging. Usually, if one clone has a distinctive sequence at one end (say, its 38' end) and another the exact same sequence at the other (58') end, it is assumed that these sequences overlap in the genome. But for *P. falciparum*, so many clone ends were AT-rich that it was difficult to assign overlaps. As a result, new stratagems had to be devised for ordering many of the chromosomal pieces, including a heavy reliance on genetic and physical 'maps' of genomic landmarks.

For example, one type of physical map used was an 'optical' map. Here, a purified chromosome is cut into segments with an enzyme known to cut DNA at particular sequences, and the segments are separated according to size by gel electrophoresis, producing the optical map. Meanwhile, the postulated sequence is 'virtually' fragmented in a computer by breaking it at the theoretical sites at which the chosen enzyme cuts. The hypothetical fragments are then sorted by length, generating a virtual map. The agreement between the optical and virtual maps for most chromosomes was reassuringly good³.

Identifying the genes

Extreme AT-richness aside, finding the genes in any eukaryotic genome can be problematic, because the protein-coding parts of genes (exons) are interrupted by non-coding regions (introns). (Eukaryotes can be loosely defined as organisms whose cells have nuclei and cytoskeletons, distinguishing them from the Bacteria and the Archaea — neither of which has introns in their coding sequences.) Although computer programs can identify the ends of exons that need to be joined together to form mature gene products, these are seldom 100% accurate. So there is often a need for validation that is not required in bacterial gene analysis. Such confirmation can be provided by studies of complementary DNA sequences, which directly reflect the mature gene products, or by identification of the encoded proteins themselves^{6,7}. To achieve the latter, the consortium used ultra-sensitive mass spectrometry — the application of which is almost sure to become a standard component of future genome projects of this sort.

Frustratingly, possible functions for fully 60% of the postulated 5,279 genes remain unknown, because these

This week's News Features section has two further articles describing reaction to publication of the *Plasmodium* genomes, and discussion of the prospects for malaria control. See pages 426 and 429.

genes match no other sequences in existing data banks. Another 5% of the genes are also classified as 'hypothetical' in this sense, although they do have counterparts — themselves with unknown functions — in other organisms. This is both surprising and disappointing. But we can be sure that many of these genes really do exist. For instance, mass spectrometry identified authentic peptides corresponding to proteins encoded by 2,391 of the genes, including many of those for which functions have not yet been found^{6,7}.

A second *Plasmodium* sequence

Remarkably, the consortium also sequenced a second plasmodial genome — that of *P. yoelii yoelii*⁸, the cause of a malaria-like disease in wild African rats. More than 60% of the *P. falciparum* genes, including most general housekeeping genes, had close relatives in the *P. yoelii yoelii* genome. But the comparison also revealed a treasure-trove of differences and rearrangements. Many of these are near the ends of chromosomes, in regions that somehow control the impressive ability of plasmodial parasites to change and thereby evade recognition by the host immune system. There is evidence in both species for the kinds of genetic rearrangements and chromosomal exchanges that might be allied to this ability. But none of the genes known to be involved in immune evasion by *P. falciparum* can be recognized in *P. yoelii yoelii*. The hope had been that comparison would reveal host-specific adaptations that could be exploited in some way, but the extreme differences have confounded that strategy.

The hunt for vulnerability

Having the entire inventory of genes for *P. falciparum* provides a complete map of its metabolic pathways, the genes encoding metabolic enzymes all being recognized by comparison with other organisms. Not only can the pathways active at different stages of the parasite life cycle be delineated, but key points at which the organism's metabolism is known to be vulnerable to attack can be seen in an overall context. For example, quinine — the first and most successful antimalarial drug — acts within a subcellular compartment of the parasite, the food vacuole, in which host haemoglobin is degraded as a foodstuff. Sulphonamides are effective because *P. falciparum* makes its own folic acid vitamins by a scheme involving *p*-aminobenzoic acid — a structure mimicked by sulphanilamide. At least four other known targets of antimalarial drugs reside in the apicoplast, an exotic quadruple-membraned compartment.

Most of these sites of drug action were known well before the genome-inspired metabolic map, although it is reassuring to see the whole landscape. It is the potential for

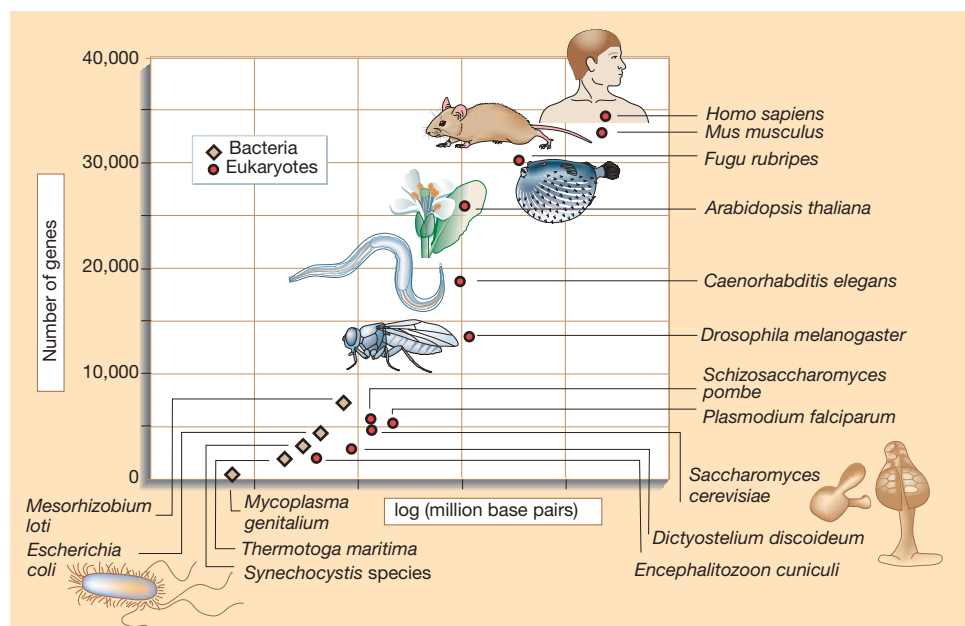


Figure 1 Some of the genomes sequenced so far. The figure shows the number of genes plotted against genome size for the 12 fully sequenced genomes of eukaryotes and a representative set of bacteria. Note the log scale for genome size, expressed as millions of base pairs.

choosing new drug targets that is exciting, however, and the consortium has now pinpointed five. For example, within the food vacuole there are several protein-degrading enzymes that might conceivably be blocked by specific inhibitors. How long it will be before these hopes are realized is unknown, but there are almost too many options to pursue. The question also arises of whether drugs that might be forthcoming would be affordable by those most in need.

Weighing the costs

During the course of this project there has been a spirited debate as to whether malaria is better attacked by large-scale genome projects¹⁰ or by more traditional public-health measures¹¹. The politically correct chorus of response has been that of course both approaches must be undertaken, especially as the true benefits of the genome studies are “down the line”^{12,13}.

Initially, the cost of the current project was put in the neighbourhood of US\$15 million¹⁴. These are not easy estimates to make because funding usually comes from multiple sources and sometimes involves third parties. To the initial figure should now be added the cost of many parallel projects, including the now published sequence¹⁵ of *Anopheles gambiae*, the mosquito that transmits malaria. The Sanger Centre and TIGR websites also list half a dozen other protozoan parasite genomes under study, including two more plasmodial strains.

So is it worth it from a medical point of view? That really remains to be seen. If one assesses what has been learned so far from the whole-genome projects (see Fig. 1) of the past six or seven years, it is clear that the

‘bio’ part of the biomedical enterprise has been the clear winner. Whether one studies molecular evolution or gene transcription, population genetics or developmental biology, cellular mechanics or signal transduction, whole-genome information is what defines the playing field. But for the most part, the promised medical benefits have been slow to materialize. Translating all of this information into new treatments and cures is not a trivial process.

That malaria was near eradication decades ago in some areas as a result of DDT spraying is more than a cruel irony^{11,16}. Indeed, since the use of that insecticide was sharply curtailed in the 1970s, 30 million people may have died from *Plasmodium*-infected mosquito bites, and ten times that number have suffered this debilitating disease. As the participants in the current study acknowledge², “genome sequences alone provide little relief to those suffering from malaria”.

Russell F. Doolittle is at the Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093-0634, USA.
e-mail: rdoolittle@ucsd.edu

1. Malaria Insight *Nature* **415**, 669–715 (2002).
2. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
3. Hall, N. *et al.* *Nature* **419**, 527–531 (2002).
4. Gardner, M. J. *et al.* *Nature* **419**, 531–534 (2002).
5. Hyman, R. W. *et al.* *Nature* **419**, 534–537 (2002).
6. Florens, L. *et al.* *Nature* **419**, 520–526 (2002).
7. Lasonder, E. *et al.* *Nature* **419**, 537–542 (2002).
8. Carlton, J. M. *et al.* *Nature* **419**, 512–519 (2002).
9. Adams, M. D. *et al.* *Science* **287**, 2185–2195 (2000).
10. Hoffmann, S. L. *Science* **290**, 1509 (2000).
11. Curtis, C. F. *Science* **290**, 1508 (2000).
12. James, A. A. *Science* **291**, 435–436 (2001).
13. Morel, C. M. *Science* **291**, 435–436 (2001).
14. Hoffmann, S. L. *et al.* *Nature* **387**, 647 (1997).
15. Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
16. Jukes, T. H. *Am. Sci.* **51**, 355–361 (1965).

The parasite genome

Biological revelations

Dyann F. Wirth

The genome of the malaria parasite was sequenced with the aim of learning more about how the parasite works, and with the hope that this would reveal potential drug targets. Has that hope been realized?

Malaria has confounded some of the best minds of the past century. A hundred years after the discovery that mosquitoes transmit *Plasmodium falciparum*, the major parasite that causes human malaria, we still do not know enough about the disease to defeat it permanently. But the papers on pages 498–542 of this issue^{1–7}, describing the complete genome sequence of *P. falciparum*, may eventually lead to new drugs and vaccines, and will certainly be an invaluable guide to future research. These papers are a testament to the success of a six-year project undertaken by an international consortium of labs and funding agencies.

Why genome sequencing?

First, a bit of background. The malaria parasite leads a complicated life (Fig. 1), existing mainly inside liver cells and red blood cells in its human host and, when residing in mosquitoes (notably *Anopheles gambiae*), being associated with the insect's gut and salivary glands. It undergoes several transformations along the way. The stages of its life cycle were

originally described more than 100 years ago and were given names based on morphology, such as merozoite, trophozoite and gametocyte (in humans), and zygote, ookinete and sporozoite (in mosquitoes). One of the most curious features of the human stages is the human immune response — there is much immune activity, but this does not control the infection effectively, nor afford protection against future infections.

Despite massive efforts to eradicate the disease in the 1950s and early 1960s, more people are infected with malaria in Africa today than at any other time in history. Over 500 million people are infected with the disease worldwide, and one-quarter of the population is at risk of infection. More than a million children die of malaria each year, mostly in Africa. And those individuals who survive suffer a combination of anaemia and immune suppression that leaves them vulnerable to other fatal illnesses. Alarming, drug resistance in the parasite is now widespread.

These stark facts emphasize the need to find new treatments for the disease and new

ways of preventing it. The genome project described in this issue^{1–7} was conceived with these goals in mind. With the wealth of information now available at the click of a mouse, malaria researchers have an unprecedented opportunity to find genes that are potentially unique to, or at least substantially different in, *P. falciparum* compared with other species; such genes may make good drug targets, with less risk of side effects.

Even before the whole genome had been sequenced, new drug targets were being identified from searches of the partially assembled sequence data for unique genes⁸. But the total sequence will provide a more complete picture of the parasite's inner workings and the chance to identify vulnerable aspects. So just what have we learnt about the parasite's biology from this package of papers, which comprises its genome sequence^{1,4–6}; a comparison of its genome with that of a rodent malaria parasite, *P. yoelii yoelii*²; and two proteomics studies of the proteins expressed at different stages in the parasite's life cycle^{3,7}? Where are the potential weaknesses? And what have we discovered about the parasite's means of evading the human immune response?

Metabolism

One notable feature of the parasite's genome¹ is the apparent absence of genes for proteins that, in other species, are key to metabolism and the energetics of mitochondria — cellular powerhouses, which produce the energy-storing molecule ATP. For example, the consortium found no predicted genes for two protein components of ATP synthase, a mitochondrial ATP-producing enzyme. (At present, many of the genes are only 'predicted': they have been identified by gene-searching algorithms, but have not yet been confirmed as bona fide genes.) Similarly, there are apparently no genes for components of a conventional NADH dehydrogenase complex, another key mitochondrial enzyme. Perhaps *P. falciparum* generates and stores energy by using novel proteins or mechanisms — potential drug targets. That the mitochondria are active, at least in sporozoites and gametocytes, seems likely, given that the proteomics analyses^{3,7} detected fragments of enzymes involved in some typical mitochondrial processes, including the tricarboxylic-acid cycle and oxidative phosphorylation.

Also interesting is the number of predicted genes — some 10% — that encode proteins associated with the apicoplast¹. This essential cellular compartment is known to be important for the biosynthesis of fatty acids and isoprenoids, components of many membrane proteins, and for iron metabolism. But analysis of these genes should reveal other possible functions, and so new drug targets. The genome sequence also identifies the molecules within the apicoplast

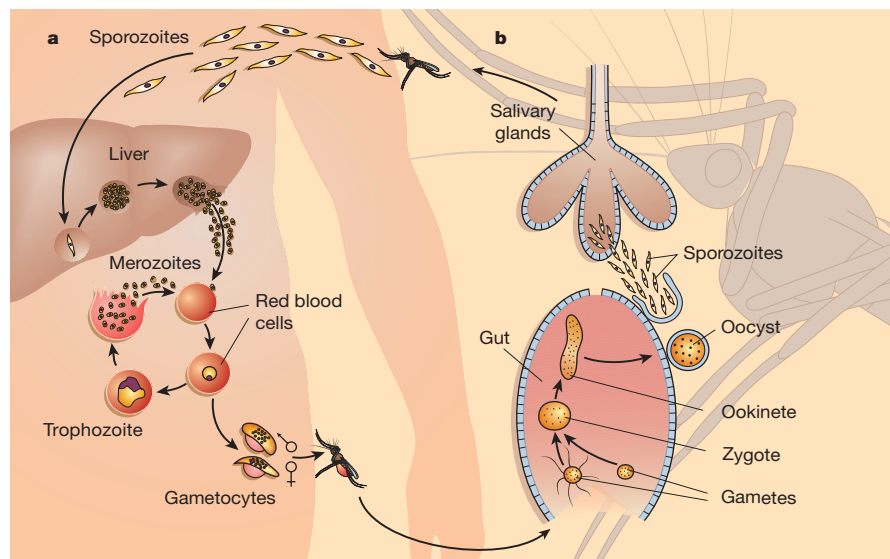


Figure 1 Life cycle of the parasite *Plasmodium falciparum*. a, When a parasite-infected mosquito feeds on a human, it injects the parasites in their sporozoite form. These travel to the liver, where they develop through several stages, finally producing merozoites which invade and multiply, via the trophozoite stage, in red blood cells. Eventually, up to 10% of all red cells become infected. (Clinical features of malaria, including fever and chills, anaemia and cerebral malaria, are all associated with infected red blood cells, and most current drugs target this stage of the life cycle.) The merozoites in a subset of infected red blood cells then develop into gametocytes. b, When another mosquito bites the infected human, it takes up blood containing gametocytes, which develop into male and female reproductive cells (gametes). These fuse in the insect's gut to form a zygote. The zygote in turn develops into the ookinete, which crosses the wall of the gut and forms a sporozoite-filled oocyst. When the oocyst bursts, the sporozoites move to the mosquito's salivary glands, and the process begins again.

that are the targets of several existing drugs⁹.

The complex life cycle of *P. falciparum* means that the parasite has had to adapt to several different environments. So it is also intriguing that, compared with the genome of the free-living budding yeast, the parasite genome¹ encodes a limited number of predicted transporter proteins for the active uptake of nutrients from the environment. In fact, entire classes of transporters seem to be missing. It may be that several genes in this class have been overlooked because they are made up of many small coding regions, which can be missed by gene-prediction algorithms. But, taken at face value, this surprising finding implies that adequate amounts of nutrients recognized by the transporters must be present at all stages of the parasite life cycle, so that there is no selective advantage in having many transporters with differing substrate specificities. Alternatively, the parasite may use previously identified pores or channels to acquire nutrients^{10,11}.

Regulating protein levels

During its life cycle, *P. falciparum* undergoes several developmental changes. One of the most dramatic is sexual differentiation and the formation of gametes, male and female reproductive cells. The proteomics studies^{3,7} of these stages have coincidentally shed light on a fundamental question: how does the parasite regulate the levels of its proteins? The genome¹ encodes relatively few predicted proteins that control the transcription of genes into messenger RNAs (the first step in making a protein). Moreover, there seem to be few transcriptional regulatory elements in the genome — or at least, there are few elements that are known from other organisms. Yet the proteomics analyses and previous studies show that protein abundance is tightly regulated.

The proteomics studies also show that proteins involved in processing mRNAs and in protein synthesis (translation) are expressed at higher levels in gametocytes, particularly female gametocytes, than in other stages. Interestingly, proteins that are present in early zygotes — which are produced from gametocytes — seem to be absent in gametocytes, although the mRNAs encoding these proteins are abundantly present. All of this is consistent with the proposal¹² that the regulation of protein levels is controlled through mRNA processing and translation, rather than by gene transcription. Perhaps this is a general feature of the parasite — another potential drug target.

In addition, one of the proteomics studies³ reveals groups of genes whose regulation appears to be coordinated. Some simultaneously expressed genes are clustered in the genome; comparison of these genes and their flanking sequences may provide further insight into how they are regulated.

Immune evasion

Arguably the most striking features of the *P. falciparum* genome are the regions near the ends of each chromosome¹. This is where families of genes that encode surface proteins, such as the *var* genes, are found. These proteins, or antigens, can sometimes be recognized by and thus stimulate the human immune system. But they have a great capacity for change, which occurs partly through the exchange of material between chromosome ends. As the genome sequence shows, the very ends of the chromosomes — the telomeres — have a complex arrangement of sequences that may facilitate such exchange (as described in ref. 13) and thereby lead to immune evasion.

The general structure of the chromosome ends is similar to that in the rodent parasite *P. yoelii yoelii*². But, surprisingly, the genes that encode the variant surface antigens in *P. falciparum* are not found in *P. yoelii yoelii*, which has a different family of variant genes, originally described in a less virulent human parasite, *P. vivax*¹⁴. This is interesting, because it suggests that *P. yoelii yoelii*, which is often used as a model of *P. falciparum*, is in some respects more similar to *P. vivax*. It is tempting to speculate that, despite their dissimilar sequences, the genes at the ends of the *P. falciparum* and *P. yoelii yoelii* chromosomes have similar functions. But that remains to be seen.

Finally, research on the *P. falciparum* *var* genes has focused on their role in enabling infected red blood cells to stick to small blood vessels in the brain. This feature is associated with the fatal form of the disease, cerebral malaria. So it is interesting that one of the proteomics analyses³ reveals that the peptides derived from many of the *var* genes occur in sporozoites, which are produced in mosquitoes and invade the human liver during the initial infection. These results point to possible alternative functions for *var* gene products.

The mosquito genome

The post-genomic era opens

Ennio De Gregorio and Bruno Lemaitre

The mosquito *Anopheles gambiae* is the main agent in the transmission of human malaria. Its genome sequence will in time help to devise control strategies, but will be a more immediate boon for insect biologists.

The papers that appear in this issue, describing the genome of the human malaria parasite *Plasmodium falciparum*, are published simultaneously with others in *Science* tackling the genome of the mosquito *Anopheles gambiae*. The connection is obvious: the parasite requires a mosquito to complete its complex life cycle and for transmission from one host to another. These two species are respectively the major parasite causing malaria and the major vector.

The complete picture

One of the most exciting aspects of this huge undertaking is that it can be related to other work. We now have the genome of the mosquito *A. gambiae*¹⁵, together with draft sequences of the human genome^{16,17}, and so can get a better handle on the interactions among three species that have long been evolving together. It is well known that certain variations in human genes are associated with a reduced susceptibility to malaria, and analysis of different human populations will no doubt reveal more on this. A close look at the mosquito genome should provide similar insights. Study of the parasite genome will reveal much about how *P. falciparum* interacts with its host and carrier, and more about the genes involved in parasite recognition by the human immune system. Decoding the information in these genomes, and translating it into effective remedies, is both a challenge and an opportunity for the scientific community. ■

Dyann F. Wirth is in the Department of Immunology and Infectious Disease, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115-6021, USA. e-mail: dfwirth@hsph.harvard.edu

1. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
2. Carlton, J. M. *et al.* *Nature* **419**, 512–519 (2002).
3. Florens, L. *et al.* *Nature* **419**, 520–526 (2002).
4. Hall, N. *et al.* *Nature* **419**, 527–531 (2002).
5. Gardner, M. J. *et al.* *Nature* **419**, 531–534 (2002).
6. Hyman, R. W. *et al.* *Nature* **419**, 534–537 (2002).
7. Lasender, E. *et al.* *Nature* **419**, 537–542 (2002).
8. Jomaa, H. *et al.* *Science* **285**, 1573–1576 (1999).
9. Waller, R. F. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 12352–12357 (1998).
10. Desai, S. A., Bezrukov, S. M. & Zimmerberg, J. *Nature* **406**, 1001–1005 (2000).
11. Kirk, K. *Nature* **406**, 949–951 (2000).
12. Dechering, K. J. *et al.* *Mol. Cell. Biol.* **19**, 967–978 (1999).
13. Freitas-Junior, L. H. *et al.* *Nature* **407**, 1018–1022 (2000).
14. del Portillo, H. A. *et al.* *Nature* **410**, 839–842 (2001).
15. Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
16. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
17. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).



Figure 1 The mosquito and the fruitfly in typical pose — *Anopheles* (top) on human skin, *Drosophila* on a banana.

has been to block parasite transmission by mosquitoes. These approaches will clearly benefit from the improved understanding of mosquito biology and mosquito interactions with *P. falciparum* that the genome sequences will make possible.

The *A. gambiae* genome¹ was sequenced by a collaboration between Celera Genomics, the French National Sequencing Centre (Genoscope) and The Institute for Genomics Research (TIGR), in association with several university laboratories. These groups used the same 'shotgun' strategy as that applied for sequencing the human, mouse and fruitfly (*Drosophila melanogaster*) genomes. Random fragments of genomic DNA were first cloned in bacteria, and sequenced, and the overlapping clones were then assembled into contiguous sequences. Unexpectedly, the high levels of genetic variation (polymorphisms) in the reference strain of *A. gambiae* used for sequencing — the PEST strain — made the genomic assembly step difficult. The genetic variation might be explained by the fact that two distinct populations of *A. gambiae* have contributed to the PEST strain, thereby creating a mosaic genome structure. This unprecedented situation required the development of new sequence-assembly strategies, and these will be a considerable asset for future genome projects — as with

mosquitoes, not all organisms are available as inbred laboratory strains.

Comparison with the fruitfly

Much of the interest in the *A. gambiae* genome will centre on comparisons with that of *D. melanogaster*, which was published two years ago². These two insects belong to the same taxonomic order, the Diptera, but inhabit distinct environments and have different lifestyles (Fig. 1). *Drosophila melanogaster* feeds on decaying organic matter, such as damaged or rotting fruit, where it also completes its life cycle, whereas *A. gambiae* feeds on sugar nectar and on the blood of vertebrate hosts. Blood meals are required for female mosquitoes to produce eggs; these are laid in water, where larvae develop and hatch. Blood feeding exposes the insect to viruses and parasites — like *Plasmodium*, these other pathogens exploit *Anopheles* as a vector for transmission.

One of the main differences between the two species is that, at 278 million base pairs, the *A. gambiae* genome is much bigger than that of *D. melanogaster* (estimated to be 180 million base pairs). But this difference is not reflected in the total number of genes, which, with 13,000–14,000 genes so far identified in both insects, is surprisingly similar. It seems that, in the course of evolution, *Drosophila* has experienced a progressive reduction both in the regions between genes and in the introns, the non-protein-coding stretches of DNA within genes.

Comparison of the coding sequences reveals that the genomes of *Anopheles* and *Drosophila* are less similar than would be expected for two species that diverged 'only' 250 million years ago. Only half of the genes in the two genomes can be interpreted as orthologues — genes in different species that have common ancestry, although their functions may differ. *Anopheles* and *Drosophila* orthologues show an average of about 56% identity in DNA sequence. As Zdobnov *et al.* point out in another of the papers in *Science*³, from the sequence standpoint, the two species differ more than do humans and pufferfish — species that diverged 450 million years ago. Some of the protein families present in both mosquito and fruitfly appear to have evolved from a common ancestral gene through independent gene-duplication in each species. The *Anopheles* genome shows several cases of such expansion which might reflect adaptation to its lifestyle. An example is the family of fibrinogen-like proteins (of which there are 58 in *Anopheles* and 13 in *Drosophila*), which in the mosquito are probably used as anticoagulant for the ingested blood meals.

Defence mechanisms

Insects have efficient immune systems for combating the various pathogens they encounter, and most of our knowledge in

this area comes from genetic and molecular studies in *Drosophila*. Finding out how *Anopheles* responds to *Plasmodium* infection is essential for obtaining clues to controlling malaria. Christophides *et al.*⁴ analysed the gene families in *A. gambiae* that are linked to insect immunity, and show that they diverge widely from those in *Drosophila*. Good examples are the prophenoloxidase enzymes (nine in the mosquito, three in the fruitfly); these enzymes catalyse the synthesis of melanin, which is associated with several defence reactions in insects.

The study by Christophides *et al.* suggests that *Anopheles* employs the same general defence mechanisms as *Drosophila*, and uses similar pathogen-activated signal-transduction pathways, but that it has adapted recognition and effector immune genes to different types of aggressors. The best characterized effector system in insects consists of antimicrobial peptides, which display a wide spectrum of antibiotic activities. Interestingly, out of seven families of these peptides found in *Drosophila*, only two are also evident in *Anopheles*: five, then, are specific to *Drosophila*. Conversely, at least one mosquito-specific antimicrobial peptide has already been identified and others might be discovered by functional studies in the future. The expression profiles of some *A. gambiae* immune genes also suggest that, like the fruitfly, the mosquito mounts specific immune responses adapted to different types of pathogen^{4,5}.

The availability of the entire DNA sequence, together with tools such as DNA microarrays and targeted gene disruption^{6–8}, will make *Anopheles* a powerful model system for studying insect biology. The genomic data will also help in developing strategies to combat malaria and other mosquito-borne human diseases, for example yellow fever, dengue, filariasis and encephalitis. Such strategies will include reducing the number and lifespan of infectious mosquitoes, analysing what attracts them to their human targets, and limiting the capacity of parasites to develop within the insect vector. Malaria is characterized by a highly complex set of interactions between the parasite, the vector and the host. Now that the genomes of all three players have been fully sequenced, the post-genomic era in combating this dreadful disease can really begin. ■

Ennio De Gregorio and Bruno Lemaitre are at the Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette, France.
e-mails: gregorio@cgm.cnrs-gif.fr
lemaitre@cgm.cnrs-gif.fr

- Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
- Special section on the *Drosophila* genome *Science* **287**, 2181–2224 (2000).
- Zdobnov, A. M. *et al.* *Science* **298**, 149–159 (2002).
- Christophides, G. K. *et al.* *Science* **298**, 159–165 (2002).
- Dimopoulos, G. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 8814–8819 (2002).
- Catteruccia, F. *et al.* *Nature* **405**, 959 (2000).
- Grossman, G. L. *et al.* *Insect Mol. Biol.* **10**, 597–604 (2001).
- Blandin, S. *et al.* *EMBO Rep.* **3**, 852–856 (2002).

Genome sequence of the human malaria parasite *Plasmodium falciparum*

Malcolm J. Gardner¹, Neil Hall², Eula Fung³, Owen White¹, Matthew Berriman², Richard W. Hyman³, Jane M. Carlton¹, Arnab Pain², Karen E. Nelson¹, Sharen Bowman^{2*}, Ian T. Paulsen¹, Keith James², Jonathan A. Eisen¹, Kim Rutherford², Steven L. Salzberg¹, Alister Craig⁴, Sue Kyes⁵, Man-Suen Chan⁵, Vishvanath Nene¹, Shamira J. Shallom¹, Bernard Suh¹, Jeremy Peterson¹, Sam Angiuoli¹, Mihaela Pertea¹, Jonathan Allen¹, Jeremy Selengut¹, Daniel Haft¹, Michael W. Mather⁶, Akhil B. Vaidya⁶, David M. A. Martin⁷, Alan H. Fairlamb⁷, Martin J. Fraunholz⁸, David S. Roos⁸, Stuart A. Ralph⁹, Geoffrey I. McFadden⁹, Leda M. Cummings¹, G. Mani Subramanian¹⁰, Chris Mungall¹¹, J. Craig Venter¹², Daniel J. Carucci¹³, Stephen L. Hoffman^{13*}, Chris Newbold⁵, Ronald W. Davis³, Claire M. Fraser¹ & Bart Barrell²

The parasite *Plasmodium falciparum* is responsible for hundreds of millions of cases of malaria, and kills more than one million African children annually. Here we report an analysis of the genome sequence of *P. falciparum* clone 3D7. The 23-megabase nuclear genome consists of 14 chromosomes, encodes about 5,300 genes, and is the most (A + T)-rich genome sequenced to date. Genes involved in antigenic variation are concentrated in the subtelomeric regions of the chromosomes. Compared to the genomes of free-living eukaryotic microbes, the genome of this intracellular parasite encodes fewer enzymes and transporters, but a large proportion of genes are devoted to immune evasion and host-parasite interactions. Many nuclear-encoded proteins are targeted to the apicoplast, an organelle involved in fatty-acid and isoprenoid metabolism. The genome sequence provides the foundation for future studies of this organism, and is being exploited in the search for new drugs and vaccines to fight malaria.

Despite more than a century of efforts to eradicate or control malaria, the disease remains a major and growing threat to the public health and economic development of countries in the tropical and subtropical regions of the world. Approximately 40% of the world's population lives in areas where malaria is transmitted. There are an estimated 300–500 million cases and up to 2.7 million deaths from malaria each year. The mortality levels are greatest in sub-Saharan Africa, where children under 5 years of age account for 90% of all deaths due to malaria¹. Human malaria is caused by infection with intracellular parasites of the genus *Plasmodium* that are transmitted by *Anopheles* mosquitoes. Of the four species of *Plasmodium* that infect humans, *Plasmodium falciparum* is the most lethal. Resistance to anti-malarial drugs and insecticides, the decay of public health infrastructure, population movements, political unrest, and environmental changes are contributing to the spread of malaria². In countries with endemic malaria, the annual economic growth rates over a 25-year period were 1.5% lower than in other countries. This implies that the cumulative effect of the lower annual economic output in a malaria-endemic country was a 50% reduction in the per capita GDP compared to a non-malarious country³. Recent studies suggest that the number of malaria cases may double in 20 years if new methods of control are not devised and implemented¹.

An international effort⁴ was launched in 1996 to sequence the *P. falciparum* genome with the expectation that the genome sequence would open new avenues for research. The sequences of two of the 14 chromosomes, representing 8% of the nuclear genome, were published previously^{5,6} and the accompanying Letters in this issue describe the sequences of chromosomes 1, 3–9 and 13 (ref. 7), 2, 10, 11 and 14 (ref. 8), and 12 (ref. 9). Here we report an analysis of the genome sequence of *P. falciparum* clone 3D7, including descriptions of chromosome structure, gene content,

functional classification of proteins, metabolism and transport, and other features of parasite biology.

Sequencing strategy

A whole chromosome shotgun sequencing strategy was used to determine the genome sequence of *P. falciparum* clone 3D7. This approach was taken because a whole genome shotgun strategy was not feasible or cost-effective with the technology that was available at the beginning of the project. Also, high-quality large insert libraries of (A + T)-rich *P. falciparum* DNA have never been constructed in *Escherichia coli*, which ruled out a clone-by-clone sequencing strategy. The chromosomes were separated on pulsed field gels, and chromosomal DNA was extracted and used to construct shotgun libraries of 1–3-kilobase (kb) fragments of sheared DNA. Eleven of the fourteen chromosomes could be resolved on the gels, but chromosomes 6, 7 and 8 could not be resolved and were sequenced as a group. The shotgun sequences were assembled into contiguous DNA sequences (contigs), in some cases with low coverage shotgun sequences of yeast artificial chromosome (YAC) clones to assist in the ordering of contigs for closure. Sequence tagged sites (STSs)¹⁰, microsatellite markers^{11,12} and HAPPY mapping⁷ were also used to place and orient contigs during the gap closure process. The high (A + T) content of the genome made gap closure extremely difficult^{7–9}. The predicted restriction enzyme maps of the chromosome sequences were compared to optical restriction maps to verify that the chromosomes had been assembled correctly¹³. Chromosomes 1–5, 9 and 12 were closed, whereas chromosomes 6–8, 10, 11, 13 and 14 contained 3–37 gaps (most <2.5 kb) per chromosome at the beginning of genome annotation. Efforts to close the remaining gaps are continuing.

¹ The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA; ² The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; ³ Stanford Genome Technology Center, 855 California Avenue, Palo Alto, California 94304, USA; ⁴ Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK; ⁵ University of Oxford, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK; ⁶ Department of Microbiology and Immunology, Drexel University College of Medicine, 2900 Queen Lane, Philadelphia, Pennsylvania 19129, USA; ⁷ School of Life Sciences, The Wellcome Trust Biocentre, The University of Dundee, Dundee DD1 5EH, UK; ⁸ Department of Biology and Genomics Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018, USA; ⁹ Plant Cell Biology Research

Centre, School of Botany, University of Melbourne, Melbourne, VIC 3010, Australia; ¹⁰ Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850, USA; ¹¹ Department of Molecular and Cellular Biology, Berkeley Drosophila Genome Project, University of California, Berkeley, California 94720, USA; ¹² The Center for the Advancement of Genomics, 1901 Research Boulevard, 6th Floor, Rockville, Maryland 20850, USA; ¹³ Malaria Program, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, Maryland 20910-7500, USA.

*Present addresses: Syngenta, Jealott's Hill International Research Centre, Bracknell, RG42 6EY, UK (S.B.); Sanaria, 308 Argosy Drive, Gaithersburg, Maryland 20878, USA (S.L.H.).

Genome structure and content

The *P. falciparum* 3D7 nuclear genome is composed of 22.8 megabases (Mb) distributed among 14 chromosomes ranging in size from approximately 0.643 to 3.29 Mb (Fig. 1, and Supplementary Figs A–N). Thus the *P. falciparum* genome is almost twice the size of the genome of the fission yeast *Schizosaccharomyces pombe*. The overall (A + T) composition is 80.6%, and rises to ~90% in introns and intergenic regions. The structures of protein-encoding genes were predicted using several gene-finding programs and manually curated. Approximately 5,300 protein-encoding genes were identified, about the same as in *S. pombe* (Table 1, and Supplementary Table A). This suggests an average gene density in *P. falciparum* of 1 gene per 4,338 base pairs (bp), slightly higher than was found previously with chromosomes 2 and 3 (1 per 4,500 bp and 1 per 4,800 bp, respectively). The higher gene density reported here is probably the result of improved gene-finding software and larger training sets that enabled the detection of genes overlooked previously⁸. Introns were predicted in 54% of *P. falciparum* genes, a proportion roughly similar to that in *S. pombe* and *Dictyostelium discoideum*, but much higher than observed in *Saccharomyces cerevisiae* where only 5% of genes contain introns. Excluding introns, the mean length of *P. falciparum* genes was 2.3 kb, substantially larger than in the other organisms in which the average gene lengths range from 1.3 to 1.6 kb. *Plasmodium falciparum* genes showed a markedly greater proportion of genes (15.5%) longer than 4 kb compared to *S. pombe* and *S. cerevisiae* (3.0% and 3.6%, respectively). The explanation for the increased gene length in *P. falciparum* is not clear. Many of these large genes encode uncharacterized proteins that may be cytosolic proteins, as they do not possess recognizable signal peptides. No transposable elements or retrotransposons were identified.

Fifty-two per cent of the predicted gene products (2,731) were detected in cell lysates prepared from several stages of the parasite life cycle by high-resolution liquid chromatography and tandem mass spectrometry^{14,15}, including many predicted proteins with no similarity to proteins in other organisms. In addition, 49% of the genes overlapped (97% identity over at least 100 nucleotides) with expressed sequence tags (ESTs) derived from several life-cycle stages. As the proteomics and EST studies performed to date may

not represent a complete sampling of all genes expressed during the complex life cycle of the parasite, this suggests that the annotation process identified substantial portions of most genes. However, in the absence of supporting EST or protein evidence, correct prediction of the 5' ends of genes and genes with multiple small exons is challenging, and the gene models should be regarded as preliminary. Additional ESTs and full-length complementary DNA sequences¹⁶ are required for the development of better training sets for gene-finding programs and the verification of the predicted genes.

The nuclear genome contains a full set of transfer RNA (tRNA) ligase genes, and 43 tRNAs were identified to bind all codons except TGT and TGC, coding for Cys; it is possible that these tRNAs are located within the currently unsequenced regions. All codons ending in C and T appear to be read by single tRNAs with a G in the first position, which is likely to read both codons via G:U wobble. Each anticodon occurs only once except for methionine (CAT), for which there are two copies, one for translation initiation and one for internal methionines, and the glycine (CCT) anticodon, which occurs twice. An unusual tRNA resembling a selenocysteinyl-tRNA was also found. A putative selenocysteine lyase was identified, which may provide selenium for synthesis of selenoproteins. Increased growth has been observed in selenium-supplemented *Plasmodium* culture¹⁷.

In almost all other eukaryotic organisms sequenced to date, the tRNA genes exhibit extensive redundancy, the only exception being the intracellular parasite *Encephalitozoon cuniculi* which contains 44 tRNAs¹⁸. Often, the abundance of specific anticodons is correlated with the codon usage of the organism^{19,20}. This is not the case in *P. falciparum*, which exhibits minimal redundancy of tRNAs. The mitochondrial genome of *Plasmodium* is small (about 6 kb) and encodes no tRNAs, so the mitochondrion must import tRNAs^{21,22}. Through their import, cytoplasmic tRNAs may serve mitochondrial protein synthesis in a manner seen with other organisms^{23,24}. The apicoplast genome appears to encode sufficient tRNAs for protein synthesis within the organelle²⁵.

Unlike many other eukaryotes, the malaria parasite genome does not contain long tandemly repeated arrays of ribosomal RNA (rRNA) genes. Instead, *Plasmodium* parasites contain several single 18S–5.8S–28S rRNA units distributed on different chromosomes.

Table 1 *Plasmodium falciparum* nuclear genome summary and comparison to other organisms

Feature	Value				
	<i>P. falciparum</i>	<i>S. pombe</i>	<i>S. cerevisiae</i>	<i>D. discoideum</i>	<i>A. thaliana</i>
Size (bp)	22,853,764	12,462,637	12,495,682	8,100,000	115,409,949
(G + C) content (%)	19.4	36.0	38.3	22.2	34.9
No. of genes	5,268*	4,929	5,770	2,799	25,498
Mean gene length† (bp)	2,283	1,426	1,424	1,626	1,310
Gene density (bp per gene)	4,338	2,528	2,088	2,600	4,526
Per cent coding	52.6	57.5	70.5	56.3	28.8
Genes with introns (%)	53.9	43	5.0	68	79
Exons					
Number	12,674	ND	ND	6,398	132,982
No. per gene	2.39	ND	NA	2.29	5.18
(G + C) content (%)	23.7	39.6	28.0	28.0	ND
Mean length (bp)	949	ND	ND	711	170
Total length (bp)	12,028,350	ND	ND	4,548,978	33,249,250
Introns					
Number	7,406	4,730	272	3,587	107,784
(G + C) content (%)	13.5	ND	NA	13.0	ND
Mean length (bp)	178.7	81	NA	177	170
Total length (bp)	1,323,509	383,130	ND	643,899	18,055,421
Intergenic regions					
(G + C) content (%)	13.6	ND	ND	14.0	ND
Mean length (bp)	1,694	952	515	786	ND
RNAs					
No. of tRNA genes	43	174	ND	73	ND
No. of 5S rRNA genes	3	30	ND	NA	ND
No. of 5.8S, 18S and 28S rRNA units	7	200–400	ND	NA	700–800

ND, not determined; NA, not applicable. *No. of genes* for *D. discoideum* are for chromosome 2 (ref. 155) and in some cases represent extrapolations to the entire genome. Sources of data for the other organisms: *S. pombe*⁶⁵, *S. cerevisiae*¹⁵⁶, *D. discoideum*¹⁵⁵ and *A. thaliana*¹⁵⁷.

*70% of these genes matched expressed sequence tags or encoded proteins detected by proteomics analyses^{14,15}.

†Excluding introns.

The sequence encoded by a rRNA gene in one unit differs from the sequence of the corresponding rRNA in the other units. Furthermore, the expression of each rRNA unit is developmentally regulated, resulting in the expression of a different set of rRNAs at different stages of the parasite life cycle^{26,27}. It is likely that by changing the properties of its ribosomes the parasite is able to alter the rate of translation, either globally or of specific messenger RNAs (mRNAs), thereby changing the rate of cell growth or altering patterns of cell development. The two types of rRNA genes previously described in *P. falciparum* are the S-type, expressed primarily in the mosquito vector, and the A-type, expressed primarily in the human host. Seven loci encoding rRNAs were identified in the genome sequence (Fig. 1). Two copies of the S-type rRNA genes are located on chromosomes 11 and 13, and two copies of the A-type genes are located on chromosomes 5 and 7. In addition, chromosome 1 contains a third, previously uncharacterized, rRNA unit that encodes 18S and 5.8S rRNAs that are almost identical to the S-type genes on chromosomes 11 and 13, but has a significantly divergent 28S rRNA gene (65% identity to the A-type and 75% identity to the S-type). The expression profiles of these genes are unknown. Chromosome 8 also contains two unusual rRNA gene units that contain 5.8S and 28S rRNA genes but do not encode 18S rRNAs; it is not known whether these genes are functional. The sequences of the 18S and 28S rRNA genes on chromosome 7 and the 28S rRNA gene on chromosome 8 are incomplete as they reside at contig ends. The 5S rRNA is encoded by three identical tandemly arrayed genes on chromosome 14.

Chromosome structure

Plasmodium falciparum chromosomes vary considerably in length, with most of the variation occurring in the subtelomeric regions. Field isolates, even those from individuals residing in a single village²⁸, exhibit extensive size polymorphism that is thought to be due to recombination events between different parasite clones during meiosis in the mosquito²⁹. Chromosome size variation is also observed in cultures of erythrocytic parasites, but is due to chromosome breakage and healing events and not to meiotic recombination^{30,31}. Subtelomeric deletions often extend well into the chromosome, and in some cases alter the cell adhesion properties of the parasite owing to the loss of the gene(s) encoding adhesion molecules^{32,33}. Because many genes involved in antigenic variation are located in the subtelomeric regions, an understanding of subtelomere structure and functional properties is essential for the elucidation of the mechanisms underlying the generation of antigenic diversity.

The subtelomeric regions of the chromosomes display a striking degree of conservation within the genome that is probably due to promiscuous inter-chromosomal exchange of subtelomeric regions. Subtelomeric exchanges occur in other eukaryotes^{34–36}, but the regions involved are much smaller (2.5–3.0 kb) in *S. cerevisiae* (data not shown). Previous studies of *P. falciparum* telomeres^{37,38} suggested that they contained six blocks of repetitive sequences that were designated telomere-associated repetitive elements (TAREs 1–6).

Whole genome analysis reveals a larger (up to 120 kb), more complex, subtelomeric repeat structure than was observed previously. The conserved regions fall into five large subtelomeric blocks (SBs; Fig. 2). The sequences within blocks 2, 4 and 5 include many tandem repeats in addition to those described previously, as well as non-repetitive regions. Subtelomeric block 1 (SB-1, equivalent to TARE-1), contains the 7-bp telomeric repeat in a variable number of near-exact copies³⁹. SB-2 contains several sub-blocks of repeats of different sizes, including TAREs 2–5 and other sequences. The beginning of SB-2 consists of about 1,000–1,300 bp of non-repetitive sequence, followed on some chromosomes by 2.5 copies of a 164-bp repeat. This is followed by another 300 bp of non-repetitive sequence, and then 10 copies of a 135-bp repeat, the main

element of TARE-2. TARE-2 is followed by 200 bp of non-repetitive sequence, and then two copies of a highly conserved 63-bp repeat. SB-2 extends for another 6 kb that contains non-repetitive sequence as well as other tandem repeats. Only four of the 28 telomeres are missing SB-2, which always occurs immediately adjacent to SB-1. A notable feature of SB-2 is the conserved order and orientation of each repeat variant as well as the sequence homology extending throughout the block. For almost any two chromosomes that were examined, a consistently ordered series of unique, identical sequences of >30 bp that are distributed across SB-2 were identified, suggesting that SB-2 is a repeat with a complex internal structure occurring once per telomere.

SB-3 consists of the Rep20 element⁴⁰, a large block of highly variable copies of a 21-bp repeat. The tandem repeats in SB-3 occur in a random order (Fig. 2). SB-4 has not been described previously, although it does contain the previously described R-FA3 sequence⁴¹. SB-4 also includes a complex mix of short (<28-bp) tandem repeats, and a 105-bp repeat that occurs once in each subtelomere. Many telomeres contain one or more *var* (variant antigen) gene exons within this block, which appear as gaps in the alignment. In five subtelomeres, fragments of 2–4 kb from SB-4 are duplicated and inverted. SB-5 is found in half of the subtelomeres, does not contain tandem repeats, and extends up to 120 kb into some chromosomes. The arrangement and composition of the subtelomeric blocks suggests frequent recombination between the telomeres.

Centromeres have not been identified experimentally in malaria parasites. However, putative centromeres were identified by comparison of the sequences of chromosomes 2 and 3 (ref. 6). Eleven of the 14 chromosomes contained a single region of 2–3 kb with extremely high (A + T) content (>97%) and imperfect short tandem repeats, features resembling the regional *S. pombe* centromeres; the 3 chromosomes lacking such regions were incomplete.

The proteome

Of the 5,268 predicted proteins, about 60% (3,208 hypothetical proteins) did not have sufficient similarity to proteins in other organisms to justify provision of functional assignments (Table 2). This is similar to what was found previously with chromosomes 2 and 3 (refs 5, 6). Thus, almost two-thirds of the proteins appear to be unique to this organism, a proportion much higher than observed in other eukaryotes. This may be a reflection of the greater evolutionary distance between *Plasmodium* and other eukaryotes that have been sequenced, exacerbated by the reduction of sequence similarity due to the (A + T) richness of the genome. Another 257 proteins (5%) had significant similarity to hypothetical proteins in other organisms. Thirty-one per cent (1,631) of the predicted proteins had one or more transmembrane domains, and 17.3% (911) of the proteins possessed putative signal peptides or signal anchors.

The Gene Ontology (GO)⁴² database is a controlled vocabulary that describes the roles of genes and gene products in organisms. GO terms were assigned manually to 2,134 gene products (40%)

Figure 1 Schematic representation of the *P. falciparum* 3D7 genome.

Protein-encoding genes are indicated by open diamonds. All genes are depicted at the same scale regardless of their size or structure. The labels indicate the name for each gene. The rows of coloured rectangles represent, from top to bottom for each chromosome, the high-level Gene Ontology assignment for each gene in the 'biological process', 'molecular function', and 'cellular component' ontologies⁴²; the life-cycle stage(s) at which each predicted gene product has been detected by proteomics techniques^{14,15}; and *Plasmodium yoelii yoelii* genes that exhibit conserved sequence and organization with genes in *P. falciparum*, as shown by a position effect analysis. Rectangles surrounding clusters of *P. yoelii* genes indicate genes shown to be linked in the *P. y. yoelii* genome¹⁶⁵. Boxes containing coloured arrowheads at the ends of each chromosome indicate subtelomeric blocks (SBs; see text and Fig. 2).

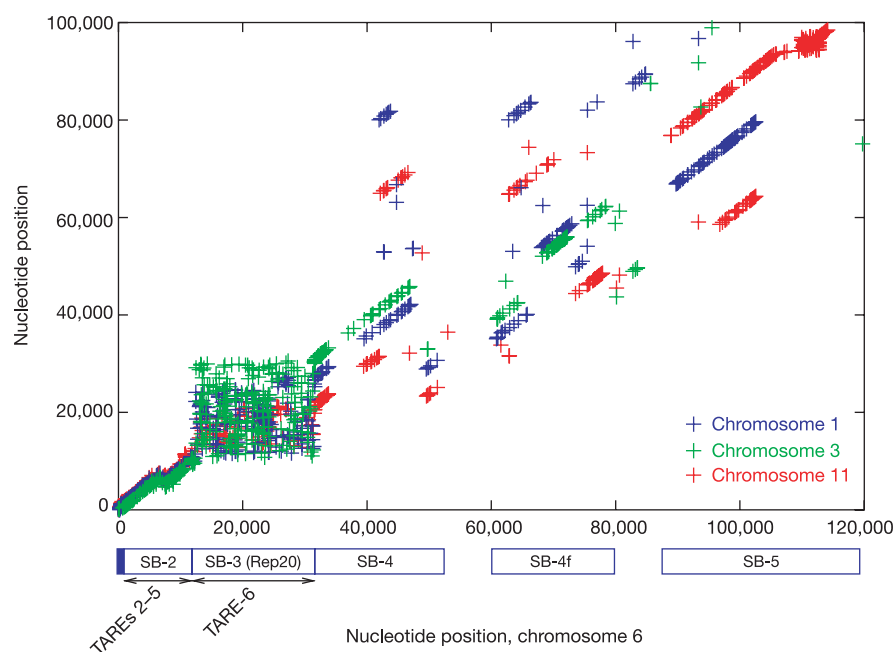


Figure 2 Alignment of subtelomeric regions of chromosomes 1, 3, 6 and 11. MUMmer2¹⁵² alignments showing exact matches between the left subtelomeric regions of chromosome 6 (horizontal axis) and chromosomes 11 (red), 1 (blue) and 3 (green), illustrating the conserved synteny between all telomeres. Each point represents an exact

and a comparison of annotation with high-level GO terms for both *S. cerevisiae* and *P. falciparum* is shown in Fig. 3. In almost all categories, higher values can be seen for *S. cerevisiae*, reflecting the greater proportion of the genome that has been characterized compared to *P. falciparum*. There are two exceptions to this pattern that reflect processes specifically connected with the parasite life cycle. At least 1.3% of *P. falciparum* genes are involved in cell-to-cell adhesion or the invasion of host cells. As discussed below (see ‘Immune evasion’), *P. falciparum* has 208 genes (3.9%) known to be involved in the evasion of the host immune system. This is reflected in the assignment of many more gene products to the GO term ‘physiological processes’ in *P. falciparum* than in *S. cerevisiae* (Fig. 3). The comparison with *S. cerevisiae* also reveals that particular

match of 40 bp or longer that is shared by two chromosomes and is not found anywhere else on either chromosome. Each collinear series of points along a diagonal represents an aligned region. SB, subtelomeric block; TARE, telomere-associated repetitive element.

categories in *P. falciparum* appear to be under-represented. Sporulation and cell budding are obvious examples (they are included in the category ‘other cell growth and/or maintenance’), but very few genes in *P. falciparum* were associated with the ‘cell organization and biogenesis’, the ‘cell cycle’, or ‘transcription factor’ categories compared to *S. cerevisiae* (Fig. 3). These differences do not necessarily imply that fewer malaria genes are involved in these processes, but highlight areas of malaria biology where knowledge is limited.

The apicoplast

Malaria parasites and other members of the phylum apicomplexa harbour a relict plastid, homologous to the chloroplasts of plants and algae^{25,43,44}. The ‘apicoplast’ is essential for parasite survival^{45,46}, but its exact role is unclear. The apicoplast is known to function in the anabolic synthesis of fatty acids^{5,47,48}, isoprenoids⁴⁹ and haeme^{50,51}, suggesting that one or more of these compounds could be exported from the apicoplast, as is known to occur in plant plastids. The apicoplast arose through a process of secondary endosymbiosis^{52–55}, in which the ancestor of all apicomplexan parasites engulfed a eukaryotic alga, and retained the algal plastid, itself the product of a prior endosymbiotic event⁵⁶. The 35-kb apicoplast genome encodes only 30 proteins²⁵, but as in mitochondria and chloroplasts, the apicoplast proteome is supplemented by proteins encoded in the nuclear genome and post-translationally targeted into the organelle by the use of a bipartite targeting signal, consisting of an amino-terminal secretory signal sequence, followed by a plastid transit peptide^{55,57–60}.

In total, 551 nuclear-encoded proteins (~10% of the predicted nuclear encoded proteins) that may be targeted to the apicoplast were identified using bioinformatic⁶¹ and laboratory-based methods. Apicoplast targeting of a few proteins has been verified by antibody localization and by the targeting of fluorescent fusion proteins to the apicoplast in transgenic *P. falciparum* or *Toxoplasma gondii*⁴⁷ parasites. Some proteins may be targeted to both the apicoplast and mitochondrion, as suggested by the observation that the total number of tRNA ligases is inadequate for independent

Table 2 The *P. falciparum* proteome

Feature	Number	Per cent
Total predicted proteins	5,268	
Hypothetical proteins	3,208	60.9
InterPro matches	2,650	52.8
Pfam matches	1,746	33.1
Gene Ontology		
Process	1,301	24.7
Function	1,244	23.6
Component	2,412	45.8
Targeted to apicoplast	551	10.4
Targeted to mitochondrion	246	4.7
Structural features		
Transmembrane domain(s)	1,631	31.0
Signal peptide	544	10.3
Signal anchor	367	7.0
Non-secretory protein	4,357	82.7

Of the apicoplast-targeted proteins, 126 were judged on the basis of experimental evidence or the predictions of multiple programs^{61,158} to be localized to the apicoplast with high confidence. Predicted apicoplast localization for 425 other proteins is based on an analysis using only one method and is of lower confidence. Predicted mitochondrial localization was based upon BLASTP searches of *S. cerevisiae* mitochondrion-targeted proteins¹⁵⁹ and TargetP¹⁵⁸ and MitoProtII¹⁶⁰ predictions; 148 genes were judged to be targeted to the mitochondrion with a high or medium confidence level, and an additional 98 genes with a lower confidence of mitochondrial targeting. Other specialized searches used the following programs and databases: InterPro¹⁶¹; Pfam¹⁶²; Gene Ontology⁴²; transmembrane domains, TMHMM¹⁶³; signal peptides and signal anchors, SignalP-2.0¹⁶⁴.

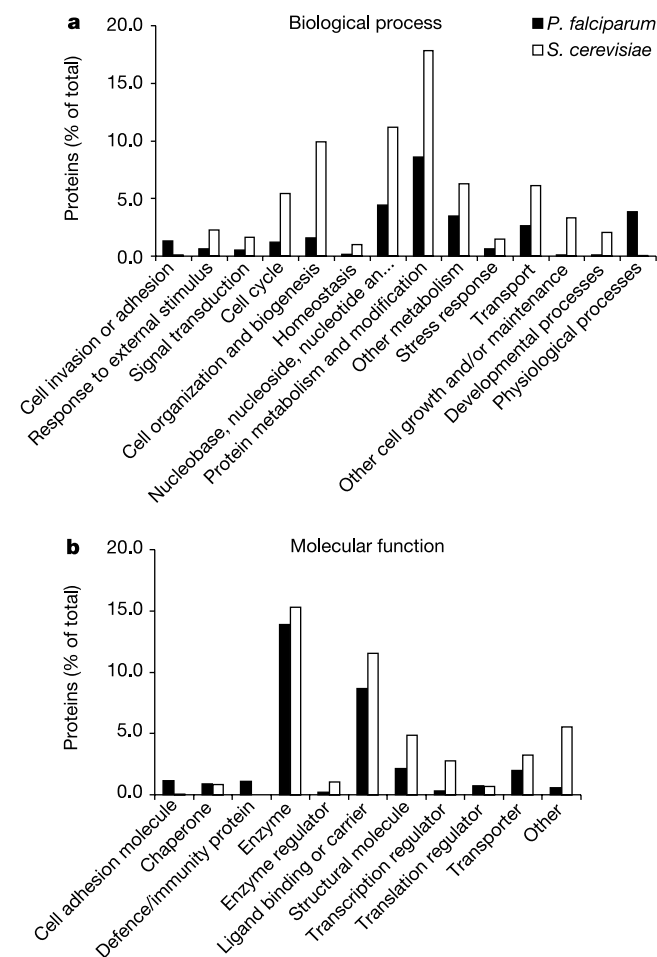


Figure 3 Gene Ontology classifications. Classification of *P. falciparum* proteins according to the 'biological process' (a) and 'molecular function' (b) ontologies of the Gene Ontology system⁴².

protein synthesis in the cytoplasm, mitochondrion and apicoplast. In plants, some proteins lack a transit peptide but are targeted to plastids via an unknown process. Proteins that use an alternative targeting pathway in *P. falciparum* would have escaped detection with the methods used.

Nuclear-encoded apicoplast proteins include housekeeping enzymes involved in DNA replication and repair, transcription, translation and post-translational modifications, cofactor synthesis, protein import, protein turnover, and specific metabolic and transport activities. No genes for photosynthesis or light perception are apparent, although ferredoxin and ferredoxin-NADP reductase are present as vestiges of photosystem I, and probably serve to recycle reducing equivalents⁶². About 60% of the putative apicoplast-targeted proteins are of unknown function. Several metabolic pathways in the organelle are distinct from host pathways and offer potential parasite-specific targets for drug therapy⁶³ (see 'Metabolism' and 'Transport' sections).

Evolution

Comparative genome analysis with other eukaryotes for which the complete genome is available (excluding the parasite *E. cuniculi*) revealed that, in terms of overall genome content, *P. falciparum* is slightly more similar to *Arabidopsis thaliana* than to other taxa. Although this is consistent with phylogenetic studies⁶⁴, it could also be due to the presence in the *P. falciparum* nuclear genome of genes derived from plastids or from the nuclear genome of the secondary endosymbiont. Thus the apparent affinity of *Plasmodium* and

Arabidopsis might not reflect the true phylogenetic history of the *P. falciparum* lineage. Comparative genomic analysis was also used to identify genes apparently duplicated in the *P. falciparum* lineage since it split from the lineages represented by the other completed genomes (Supplementary Table B).

There are 237 *P. falciparum* proteins with strong matches to proteins in all completed eukaryotic genomes but no matches to proteins, even at low stringency, in any complete prokaryotic proteome (Supplementary Table C). These proteins help to define the differences between eukaryotes and prokaryotes. Proteins in this list include those with roles in cytoskeleton construction and maintenance, chromatin packaging and modification, cell cycle regulation, intracellular signalling, transcription, translation, replication, and many proteins of unknown function. This list overlaps with, but is somewhat larger than, the list generated by an analysis of the *S. pombe* genome⁶⁵. The differences are probably due in part to the different stringencies used to identify the presence or absence of homologues in the two studies.

A large number of nuclear-encoded genes in most eukaryotic species trace their evolutionary origins to genes from organelles that have been transferred to the nucleus during the course of eukaryotic evolution. Similarity searches against other complete genomes were used to identify *P. falciparum* nuclear-encoded genes that may be derived from organellar genomes. Because similarity searches are not an ideal method for inferring evolutionary relatedness⁶⁶, phylogenetic analysis was used to gain a more accurate picture of the evolutionary history of these genes. Out of 200 candidates examined, 60 genes were identified as being of probable mitochondrial origin. The proteins encoded by these genes include many with known or expected mitochondrial functions (for example, the tricarboxylic acid (TCA) cycle, protein translation, oxidative damage protection, the synthesis of haem, ubiquinone and pyrimidines), as well as proteins of unknown function. Out of 300 candidates examined, 30 were identified as being of probable plastid origin, including genes with predicted roles in transcription and translation, protein cleavage and degradation, the synthesis of isoprenoids and fatty acids, and those encoding four subunits of the pyruvate dehydrogenase complex. The origin of many candidate organelle-derived genes could not be conclusively determined, in part due to the problems inherent in analysing genes of very high (A + T) content. Nevertheless, it appears likely that the total number of plastid-derived genes in *P. falciparum* will be significantly lower than that in the plant *A. thaliana* (estimated to be over 1,000). Phylogenetic analysis reveals that, as with the *A. thaliana* plastid, many of the genes predicted to be targeted to the apicoplast are apparently not of plastid origin. Of 333 putative apicoplast-targeted genes for which trees were constructed, only 26 could be assigned a probable plastid origin. In contrast, 35 were assigned a probable mitochondrial origin and another 85 might be of mitochondrial origin but are probably not of plastid origin (they group with eukaryotes that have not had plastids in their history, such as humans and fungi, but the relationship to mitochondrial ancestors is not clear). The apparent non-plastid origin of these genes could either be due to inaccuracies in the targeting predictions or to the co-option of genes derived from the mitochondria or the nucleus to function in the plastid, as has been shown to occur in some plant species⁶⁷.

Metabolism

Biochemical studies of the malaria parasite have been restricted primarily to the intra-erythrocytic stage of the life cycle, owing to the difficulty of obtaining suitable quantities of material from the other life-cycle stages. Analysis of the genome sequence provides a global view of the metabolic potential of *P. falciparum* irrespective of the life-cycle stage (Fig. 4). Of the 5,268 predicted proteins, 733 (~14%) were identified as enzymes, of which 435 (~8%) were assigned Enzyme Commission (EC) numbers. This is considerably

fewer than the roughly one-quarter to one-third of the genes in bacterial and archaeal genomes that can be mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway diagrams⁶⁸, or the 17% of *S. cerevisiae* open reading frames that can be assigned EC numbers. This suggests either that *P. falciparum* has a smaller proportion of its genome devoted to enzymes, or that enzymes are more difficult to identify in *P. falciparum* by sequence similarity methods. (This difficulty can be attributed either to the great evolutionary distance between *P. falciparum* and other well-studied organisms, or to the high (A + T) content of the genome.) A few genes might have escaped detection because they were located in the small regions of the genome that remain to be sequenced (Table 1). However, many biochemical pathways could be reconstructed in their entirety, suggesting that the similarity-searching approach was for the most part successful, and that the relative paucity of enzymes in *P. falciparum* may be related to its parasitic life-style. A similar

picture has emerged in the analysis of transporters (see 'Transport').

In erythrocytic stages, *P. falciparum* relies principally on anaerobic glycolysis for energy production, with regeneration of NAD⁺ by conversion of pyruvate to lactate⁶⁹. Genes encoding all of the enzymes necessary for a functional glycolytic pathway were identified, including a phosphofructokinase (PFK) that has sequence similarity to the pyrophosphate-dependent class of enzymes but which is probably ATP-dependent on the basis of the characterization of the homologous enzyme in *Plasmodium berghiei*^{70,71}. A second putative pyrophosphate-dependent PFK was also identified which possessed N- and carboxy-terminal extensions that could represent targeting sequences.

A gene encoding fructose bisphosphatase could not be detected, suggesting that gluconeogenesis is absent, as are enzymes for synthesis of trehalose, glycogen or other carbohydrate stores. Candidate genes for all but one enzyme of the conventional pentose

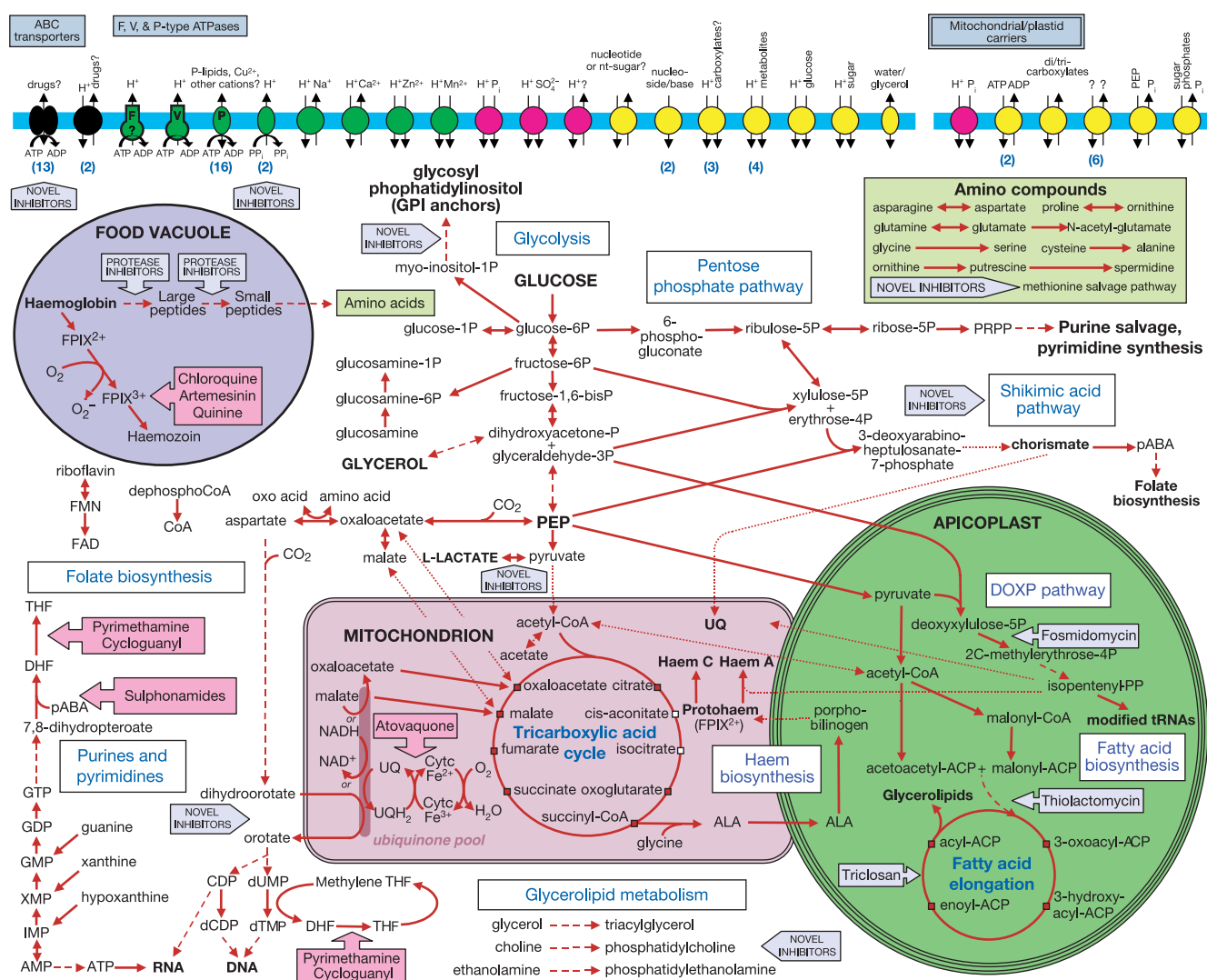


Figure 4 Overview of metabolism and transport in *P. falciparum*. Glucose and glycerol provide the major carbon sources for malaria parasites. Metabolic steps are indicated by arrows, with broken lines indicating multiple intervening steps not shown; dotted arrows indicate incomplete, unknown or questionable pathways. Known or potential organellar localization is shown for pathways associated with the food vacuole, mitochondrion and apicoplast. Small white squares indicate TCA (tricarboxylic acid) cycle metabolites that may be derived from outside the mitochondrion. Fuschia block arrows indicate the steps inhibited by antimalarials; grey block arrows highlight potential drug targets. Transporters are grouped by substrate specificity: inorganic cations (green), inorganic anions

(magenta), organic nutrients (yellow), drug efflux and other (black). Arrows indicate direction of transport for substrates (and coupling ions, where appropriate). Numbers in parentheses indicate the presence of multiple transporter genes with similar substrate predictions. Membrane transporters of unknown or putative subcellular localization are shown in a generic membrane (blue bar). Abbreviations: ACP, acyl carrier protein; ALA, aminolevulinic acid; CoA, coenzyme A; DHF, dihydrofolate; DOXP, deoxyxylulose phosphate; FPIX²⁺ and FPIX³⁺, ferro- and ferriprotoporphyrin IX, respectively; pABA, para-aminobenzoic acid; PEP, phosphoenolpyruvate; P_i, phosphate; PP_i, pyrophosphate; PRPP, phosphoribosyl pyrophosphate; THF, tetrahydrofolate; UQ, ubiquinone.

phosphate pathway were found. These include a bifunctional glucose-6-phosphate dehydrogenase/6-phosphogluconate dehydrogenase required to generate NADPH and ribose 5-phosphate for other biosynthetic pathways^{72,73}. Transaldolase appears to be absent, but erythrose 4-phosphate required for the chorismate pathway could probably be generated from the glycolytic intermediates fructose 6-phosphate and glyceraldehyde 3-phosphate via a putative transketolase (Fig. 4).

The genes necessary for a complete TCA cycle, including a complete pyruvate dehydrogenase complex, were identified. However, it remains unclear whether the TCA cycle is used for the full oxidation of products of glycolysis, or whether it is used to supply intermediates for other biosynthetic pathways. The pyruvate dehydrogenase complex seems to be localized in the apicoplast, and the only protein with significant similarity to aconitases has been reported to be a cytosolic iron-response element binding protein that did not possess aconitase activity⁷⁴. Also, malate dehydrogenase appears to be cytosolic rather than mitochondrial, even though it seems to have originated from the mitochondrial genome⁷⁵. Genes encoding malate-quinone oxidoreductase and type I fumarate hydratase are present. Malate-quinone oxidoreductase, which is probably targeted to the mitochondrion, may well replace malate dehydrogenase in the TCA cycle, as it does in *Helicobacter pylori*. A gene encoding phosphoenolpyruvate carboxylase (PEPC) was also found. Like bacteria and plants, *P. falciparum* may cope with a drain of TCA cycle intermediates by using phosphoenolpyruvate (PEP) to replenish oxaloacetate (Fig. 4). This would seem to be supported by reports of CO₂-incorporating activity in asexual stage parasite cultures⁷⁶. Thus, the TCA cycle appears to be unconventional in erythrocytic stages, and may serve mainly to synthesize succinyl-CoA, which in turn can be used in the haem biosynthesis pathway.

Genes encoding all subunits of the catalytic F₁ portion of ATP synthase, the protein that confers oligomycin sensitivity, and the gene that encodes the proteolipid subunit *c* for the F₀ portion of ATP synthase, were detected in the parasite genome. The F₀ *a* and *b* subunits could not be detected, raising the question as to whether the ATP synthase is functional. Because parts of the genome sequence are incomplete, the presence of the *a* and *b* subunits could not be ruled out. Erythrocytic parasites derive ATP through glycolysis and the mitochondrial contribution to the ATP pool in these stages appears to be minimal^{77,78}. It is possible that the ATP synthase functions in the insect or sexual stages of the parasite. However, in the absence of the F₀ *a* and *b* subunits, an ATP synthase cannot use the proton gradient⁷⁹.

A functional mitochondrion requires the generation of an electrochemical gradient across the inner membrane. But the *P. falciparum* genome seems to lack genes encoding components of a conventional NADH dehydrogenase complex I. Instead, a single subunit NADH dehydrogenase gene specifies an enzyme that can accomplish ubiquinone reduction without proton pumping, thus constituting a non-electrogenic step. Other dehydrogenases targeted to the mitochondrion also serve to reduce ubiquinone in *P. falciparum*, including dihydroorotate dehydrogenase, a critical enzyme in the essential pyrimidine biosynthesis pathway⁸⁰. The parasite genome contains some genes specifying ubiquinone synthesis enzymes, in agreement with recent metabolic labelling studies⁸¹. Re-oxidation of ubiquinol is carried out by the cytochrome *bc1* complex that transfers electrons to cytochrome *c*, and is accompanied by proton translocation⁸². Apocytochrome *b* of this complex is encoded by the mitochondrial genome^{21,22}, but the rest of the components are encoded by nuclear genes. Ubiquinol cycling is a critical step in mitochondrial physiology, and its selective inhibition by hydroxynaphthoquinones is the basis for their antimalarial action⁸³. The final step in electron transport is carried out by the proton-pumping cytochrome *c* oxidase complex, of which only two subunits are encoded in the mitochondrial DNA (mtDNA). In most eukaryotes, subunit II of cytochrome *c* oxidase is encoded by a gene on the

mitochondrial genome. In *P. falciparum*, however, the *coxII* gene is divided such that the N-terminal portion is encoded on chromosome 13 and the C-terminal portion on chromosome 14. A similar division of the *coxII* gene is also seen in the unicellular alga, *Chlamydomonas reinhardtii*⁸⁴. An alternative oxidase that transfers electrons directly from ubiquinol to oxygen has been seen in plants as well in many protists, and an earlier biochemical study suggested its presence in *P. falciparum*⁸⁵. The genome sequence, however, fails to reveal such an oxidase gene.

Biochemical, genetic and chemotherapeutic data suggest that malaria and other apicomplexan parasites synthesize chorismate from erythrose 4-phosphate and phosphoenolpyruvate via the shikimate pathway^{86–89}. It was initially suggested that the pathway was located in the apicoplast⁸⁸, but chorismate synthase is phylogenetically unrelated to plastid isoforms⁹⁰ and has subsequently been localized to the cytosol⁹¹. The genes for the preceding enzymes in the pathway could not be identified with certainty, but a BLASTP search with the *S. cerevisiae* arom polypeptide⁹², which catalyses 5 of the preceding steps, identified a protein with a low level of similarity (E value 7.9×10^{-8}).

In many organisms, chorismate is the pivotal precursor to several pathways, including the biosynthesis of aromatic amino acids and ubiquinone. We found no evidence, on the basis of similarity searches, for a role of chorismate in the synthesis of tryptophan, tyrosine or phenylalanine, although *para*-aminobenzoate (pABA) synthase does have a high degree of similarity to anthranilate (2-amino benzoate) synthase, the enzyme catalysing the first step in tryptophan synthesis from chorismate. In accordance with the supposition that the malaria parasite obtains all of its amino acids either by salvage from the host or by globin digestion, we found no enzymes required for the synthesis of other amino acids with the exception of enzymes required for glycine-serine, cysteine-alanine, aspartate-asparagine, proline-ornithine and glutamine-glutamate interconversions. In addition to pABA synthase, all but one of the enzymes (dihydroneopterin aldolase) required for *de novo* synthesis of folate from GTP were identified.

Several studies have shown that the erythrocytic stages of *P. falciparum* are incapable of *de novo* purine synthesis (reviewed in ref. 80). This statement can now be extended to all life-cycle stages, as only adenylosuccinate lyase, one of the 10 enzymes required to make inosine monophosphate (IMP) from phosphoribosyl pyrophosphate, was identified. This enzyme also plays a role in purine salvage by converting IMP to AMP. Purine transporters and enzymes for the interconversion of purine bases and nucleosides are also present. The parasite can synthesize pyrimidines *de novo* from glutamine, bicarbonate and aspartate, and the genes for each step are present. Deoxyribonucleotides are formed via an aerobic ribonucleoside diphosphate reductase^{93,94}, which is linked via thioredoxin to thioredoxin reductase. Gene knockout experiments have recently shown that thioredoxin reductase is essential for parasite survival⁹⁵.

The intraerythrocytic stages of the malaria parasite uses haemoglobin from the erythrocyte cytoplasm as a food source, hydrolysing globin to small peptides, and releasing haem that is detoxified in the form of haemozoin. Although large amounts of haem are toxic to the parasite, *de novo* haem biosynthesis has been reported⁹⁶ and presumably provides a mechanism by which the parasite can segregate host-derived haem from haem required for synthesis of its own iron-containing proteins. However, it has been unclear whether *de novo* synthesis occurs using imported host enzymes⁹⁷ or parasite-derived enzymes. Genes encoding the first two enzymes in the haem biosynthetic pathway, aminolevulinic synthase⁹⁸ and aminolevulinic dehydratase⁹⁹, were cloned previously, and genes encoding every other enzyme in the pathway except for uroporphyrinogen-III synthase were found (Fig. 4).

Haem and iron-sulphur clusters form redox prosthetic groups for a wide range of proteins, many of which are localized to the

mitochondrion and apicoplast. The parasite genome appears to encode enzymes required for the synthesis of these molecules. There are two putative cysteine desulphurase genes, one which also has homology to selenocysteine lyase and may be targeted to the mitochondrion, and the second which may be targeted to the apicoplast, suggesting organelle specific generation of elemental sulphur to be used in Fe-S cluster proteins. The subcellular localization of the enzymes involved in haem synthesis is uncertain. Ferrochelatase and two haem lyases are likely to be localized in the mitochondrion.

The role of the apicoplast in type II fatty-acid biosynthesis was described previously^{5,47}. The genes encoding all enzymes in the pathway have now been elucidated, except for a thioesterase required for chain termination. No evidence was found for the associative (type I) pathway for fatty-acid biosynthesis common to most eukaryotes. The apicoplast also houses the machinery for mevalonate-independent isoprenoid synthesis. Because it is not present in mammals, the biosynthesis of isopentyl diphosphate from pyruvate and glyceraldehyde-3-phosphate provides several attractive targets for chemotherapy. Three enzymes in the pathway have been identified, including 1-deoxy-D-xylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase⁴⁹, and 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase^{100,101}. One predicted protein was similar to the fourth enzyme, 2C-methyl-D-erythritol-4-phosphate cytidyltransferase (BLASTP E value 9.6×10^{-15}).

Transport

On the basis of genome analysis, *P. falciparum* possesses a very limited repertoire of membrane transporters, particularly for uptake of organic nutrients, compared to other sequenced eukaryotes (Fig. 5). For instance, there are only six *P. falciparum* members of the major facilitator superfamily (MFS) and one member of the amino acid/polyamine/choline APC family, less than 10% of the numbers seen in *S. cerevisiae*, *S. pombe* or *Caenorhabditis elegans* (Fig. 5). The apparent lack of solute transporters in *P. falciparum* correlates with the lower percentage of multispanning membrane proteins compared with other eukaryotic organisms (Fig. 5). The predicted transport capabilities of *P. falciparum* resemble those of obligate intracellular prokaryotic parasites, which also possess a limited complement of transporters for organic solutes¹⁰².

A complete catalogue of the identified transporters is presented in Fig. 4. In addition to the glucose/proton symporter¹⁰³ and the water/glycerol channel¹⁰⁴, one other probable sugar transporter and three carboxylate transporters were identified; one or more of the latter are probably responsible for the lactate and pyruvate/proton symport activity of *P. falciparum*¹⁰⁵. Two nucleoside/nucleobase transporters are encoded on the *P. falciparum* genome, one of which has been localized to the parasite plasma membrane¹⁰⁶. No obvious amino-acid transporters were detected, which emphasizes the importance of haemoglobin digestion within the food vacuole as an important source of amino acids for the erythrocytic stages of the parasite. How the insect stages of the parasite acquire amino acids and other important nutrients is unknown, but four metabolic uptake systems were identified whose substrate specificity could not be predicted with confidence. The parasite may also possess novel proteins that mediate these activities. Nine members of the mitochondrial carrier family are present in *P. falciparum*, including an ATP/ADP exchanger¹⁰⁷ and a di/tri-carboxylate exchanger, probably involved in transport of TCA cycle intermediates across the mitochondrial membrane. Probable phosphoenolpyruvate/phosphate and sugar phosphate/phosphate antiporters most similar to those of plant chloroplasts were identified, suggesting that these transporters are targeted to the apicoplast membrane. The former may enable uptake of phosphoenolpyruvate as a precursor of fatty-acid biosynthesis.

A more extensive set of transporters could be identified for

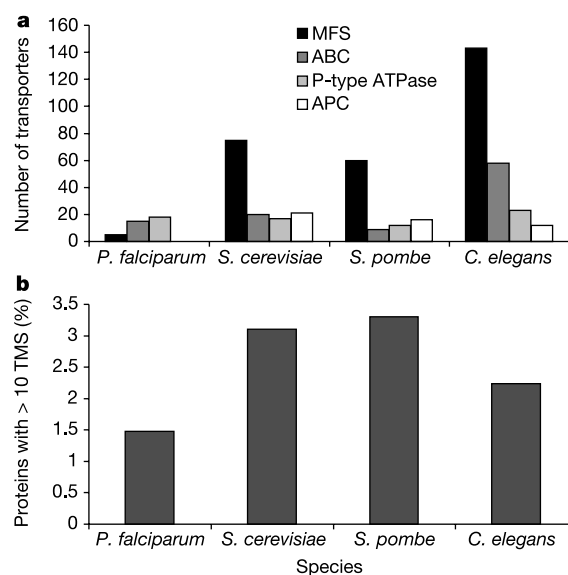


Figure 5 Analysis of transporters in *P. falciparum*. **a**, Comparison of the numbers of transporters belonging to the major facilitator superfamily (MFS), ATP-binding cassette (ABC) family, P-type ATPase family and the amino acid/polyamine/choline (APC) family in *P. falciparum* and other eukaryotes. Analyses were performed as previously described¹⁰². **b**, Comparison of the numbers of proteins with ten or more predicted transmembrane segments¹⁶³ (TMS) in *P. falciparum* and other eukaryotes. Prediction of membrane spanning segments was performed using TMHMM.

the transport of inorganic ions and for export of drugs and hydrophobic compounds. Sodium/proton and calcium/proton exchangers were identified, as well as other metal cation transporters, including a substantial set of 16 P-type ATPases. An Nramp divalent cation transporter was identified which may be specific for manganese or iron. *Plasmodium falciparum* contains all subunits of V-type ATPases as well as two proton translocating pyrophosphatases¹⁰⁸, which could be used to generate a proton motive force, possibly across the parasite plasma membrane as well as across a vacuolar membrane. The proton pumping pyrophosphatases are not present in mammals, and could form attractive antimalarial targets. Only a single copy of the *P. falciparum* chloroquine-resistance gene *crt* is present, but multiple homologues of the multidrug resistance pump *mdr1* and other predicted multidrug transporters were identified (Fig. 3). Mutations in *crt* seem to have a central role in the development of chloroquine resistance¹⁰⁹.

Plasmodium falciparum infection of erythrocytes causes a variety of pleiotropic changes in host membrane transport. Patch clamp analysis has described a novel broad-specificity channel activated or inserted in the red blood cell membrane by *P. falciparum* infection that allows uptake of various nutrients¹¹⁰. If this channel is encoded by the parasite, it is not obvious from genome analysis, because no clear homologues of eukaryotic sodium, potassium or chloride ion channels could be identified. This suggests that *P. falciparum* may use one or more novel membrane channels for this activity.

DNA replication, repair and recombination

DNA repair processes are involved in maintenance of genomic integrity in response to DNA damaging agents such as irradiation, chemicals and oxygen radicals, as well as errors in DNA metabolism such as misincorporation during DNA replication. The *P. falciparum* genome encodes at least some components of the major DNA repair processes that have been found in other eukaryotes^{111,112}. The core of eukaryotic nucleotide excision repair is present (XPB/Rad25, XPG/Rad2, XPF/Rad1, XPD/Rad3, ERCC1) although some highly conserved proteins with more accessory roles

could not be found (for example, XPA/Rad4, XPC). The same is true for homologous recombinational repair with core proteins such as MRE11, DMC1, Rad50 and Rad51 present but accessory proteins such as NBS1 and XRS2 not yet found. These accessory proteins tend to be poorly conserved and have not been found outside of animals or yeast, respectively, and thus may be either absent or difficult to identify in *P. falciparum*. However, it is interesting that Archaea possess many of the core proteins but not the accessory proteins for these repair processes, suggesting that many of the accessory eukaryotic repair proteins evolved after *P. falciparum* diverged from other eukaryotes.

The presence of MutL and MutS homologues including possible orthologues of MSH2, MSH6, MLH1 and PMS1 suggests that *P. falciparum* can perform post-replication mismatch repair. Orthologues of MSH4 and MSH5, which are involved in meiotic crossing over in other eukaryotes, are apparently absent in *P. falciparum*. The repair of at least some damaged bases may be performed by the combined action of the four base excision repair glycosylase homologues and one of the apurinic/apyrimidinic (AP) endonucleases (homologues of Xth and Nfo are present). Experimental evidence suggests that this is done by the long-patch pathway¹¹³.

The presence of a class II photolyase homologue is intriguing, because it is not clear whether *P. falciparum* is exposed to significant amounts of ultraviolet irradiation during its life cycle. It is possible that this protein functions as a blue-light receptor instead of a photolyase, as do members of this gene family in some organisms such as humans. Perhaps most interesting is the apparent absence of homologues of any of the genes encoding enzymes known to be involved in non-homologous end joining (NHEJ) in eukaryotes (for example, Ku70, Ku86, Ligase IV and XRCC1)¹¹². NHEJ is involved in the repair of double strand breaks induced by irradiation and chemicals in other eukaryotes (such as yeast and humans), and is also involved in a few cellular processes that create double strand breaks (for example, VDJ recombination in the immune system in humans). The role of NHEJ in repairing radiation-induced double strand breaks varies between species¹¹⁴. For example, in humans, cells with defects in NHEJ are highly sensitive to γ -irradiation while yeast mutants are not. Double strand breaks in yeast are repaired primarily by homologous recombination. As NHEJ is involved in regulating telomere stability in other organisms, its apparent absence in *P. falciparum* may explain some of the unusual properties of the telomeres in this species¹¹⁵.

Secretory pathway

Plasmodium falciparum contains genes encoding proteins that are important in protein transport in other eukaryotic organisms, but the organelles associated with a classical secretory pathway and protein transport are difficult to discern at an ultra-structural level¹¹⁶. In order to identify additional proteins that may have a role in protein translocation and secretion, the *P. falciparum* protein database was searched with *S. cerevisiae* proteins with GO assignments for involvement in protein export. We identified potential homologues of important components of the signal recognition particle, the translocon, the signal peptidase complex and many components that allow vesicle assembly, docking and fusion, such as COPI and COPII, clathrin, adaptin, v- and t-SNARE and GTP binding proteins. The presence of Sec62 and Sec63 orthologues raises the possibility of post-translational translocation of proteins, as found in *S. cerevisiae*.

Although *P. falciparum* contains many of the components associated with a classical secretory system and vesicular transport of proteins, the parasite secretory pathway has unusual features. The parasite develops within a parasitophorous vacuole that is formed during the invasion of the host cell, and the parasite modifies the host erythrocyte by the export of parasite-encoded proteins¹¹⁷. The mechanism(s) by which these proteins, some of which lack signal peptide sequences, are transported through and targeted beyond the

membrane of the parasitophorous vacuole remains unknown. But these mechanisms are of particular importance because many of the proteins that contribute to the development of severe disease are exported to the cytoplasm and plasma membrane of infected erythrocytes.

Attempts to resolve these observations resulted in the proposal of a secondary secretory pathway¹¹⁸. More recent studies suggest export of COPII vesicle coat proteins, Sar1 and Sec31, to the erythrocyte cytoplasm as a mechanism of inducing vesicle formation in the host cell, thereby targeting parasite proteins beyond the parasitophorous vacuole, a new model in cell biology^{119,120}. A homologue of *N*-ethylmaleimide-sensitive factor (NSF), a component of vesicular transport, has also been located to the erythrocyte cytoplasm¹²¹. The 41-2 antigen of *P. falciparum*, which is also found in the erythrocyte cytoplasm and plasma membrane¹²², is homologous with BET3, a subunit of the *S. cerevisiae* transport protein particle (TRAPP) that mediates endoplasmic reticulum to Golgi vesicle docking and fusion¹²³. It is not clear how these proteins are targeted to the cytoplasm, as they lack an obvious signal peptide. Nevertheless, the expanded list of protein-transport-associated genes identified in the *P. falciparum* genome should facilitate the development of specific probes to further elucidate the intra- and extracellular compartments of its protein transport system.

Immune evasion

In common with other organisms, highly variable gene families are clustered towards the telomeres. *Plasmodium falciparum* contains three such families termed *var*, *rif* and *stevor*, which code for proteins known as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), repetitive interspersed family (rifin) and sub-telomeric variable open reading frame (stevor), respectively^{5,124–130}. The 3D7 genome contains 59 *var*, 149 *rif* and 28 *stevor* genes, but for each family there are also a number of pseudogenes and gene truncations present.

The *var* genes code for proteins which are exported to the surface of infected red blood cells where they mediate adherence to host endothelial receptors¹³¹, resulting in the sequestration of infected cells in a variety of organs. These and other adherence properties^{132–135} are important virulence factors that contribute to the development of severe disease. Rifins, products of the *rif* genes, are also expressed on the surface of infected red cells and undergo antigenic variation¹³¹. Proteins encoded by *stevor* genes show sequence similarity to rifins, but they are less polymorphic than the rifins¹²⁹. The function of rifins and stevors is unknown. PfEMP1 proteins are targets of the host protective antibody response¹³⁶, but transcriptional switching between *var* genes permits antigenic variation and a means of immune evasion, facilitating chronic infection and transmission. Products of the *var* gene family are thus central to the pathogenesis of malaria and to the induction of protective immunity.

Figure 6 shows the genome-wide arrangement of these multigene families. In the 24 chromosomal ends that have a *var* gene as the first transcriptional unit, there are three basic types of gene arrangement. Eight have the general pattern *var-rif var* + / – (*rif/stevor*)_m, ten can be described as *var-(rif/stevor)*_m, three have a *var* gene alone and two have two or more adjacent *var* genes. This telomeric organization is consistent with exchange between chromosome ends, although the extent of this re-assortment may be limited by the varied gene combinations. The *var*, *rif* and *stevor* genes consist of two exons. The first *var* exon is between 3.5 and 9.0 kb in length, polymorphic and encodes an extracellular region of the protein. The second exon is between 1.0 and 1.5 kb, and encodes a conserved cytoplasmic tail that contains acidic amino-acid residues (ATS; 'acidic terminal sequence'). The first *rif* and *stevor* exons are about 50–75 bp in length, and encode a putative signal sequence while the second exon is about 1 kb in length, with the *rif* exon being on average slightly larger than that for *stevor*. The rifin sequences fall into two major

subgroups determined by the presence or absence of a consensus peptide sequence, KEL (X_{15}) IPTCVCR, approximately 100 amino acids from the N terminus. The *var* genes are made up of three recognizable domains known as 'Duffy binding like' (DBL); 'cysteine rich interdomain region' (CIDR) and 'constant2' (C2)^{137–139}. Alignment of sequences existing before the *P. falciparum* genome project had placed each of these domains into a number of sub-classes; α to ϵ for DBL domains, and α to γ for CIDR domains. Despite these recognizable signatures, there is a low level of sequence similarity even between domains of the same sub-type. Alignment and tree construction of the DBL domains identified here showed that a small number did not fit well into existing categories, and have been termed DBL-X. Similar analysis of all 3D7 CIDR sequences showed that with this data they were best described as CIDR α or CIDR non- α , as distinct tree branches for the other domain types were not observed. In terms of domain type and order, 16 types of *var* gene sequences were identified in this study.

Type 1 *var* genes, consisting of DBL α , CIDR α , DBL δ , and CIDR non- α followed by the ATS, are the most common structures, with 38 genes in this category (Fig. 6b). A total of 58 *var* genes commence with a DBL α domain, and in 51 cases this is followed by CIDR α , and in 46 *var* genes the last domain of the first exon is CIDR non- α . Four *var* genes are atypical with the first exon consisting solely of DBL domains (type 3 and type 13). There is non-randomness in the ordering and pairing of DBL and CIDR sub-domains¹⁴⁰, suggesting that some—for example, DBL δ –CIDR non- α and DBL β –C2

(Table 3)—should either be considered as functional–structural combinations, or that recombination in these areas is not favoured, thereby preserving the arrangement. Eighteen of the 24 telomeric proximal *var* genes are of type 1. With two exceptions, type 4 on chromosome 7 and type 9 on chromosome 11, all of the telomeric *var* genes are transcribed towards the centromere. The inverted position of the two *var* genes may hinder homologous recombination at these loci in telomeric clusters that are formed during asexual multiplication¹¹⁵. A further 12 *var* genes are located near to telomeres, with the remaining *var* genes forming internal clusters on chromosomes 4, 7, 8 and 12 and a single internal gene being located on chromosome 6.

Alignment of sequences 1.5 kb upstream of all of the *var* genes revealed three classes of sequences, upsA, upsB and upsC (of which there are 11, 35 and 13 members, respectively) that show preferential association with different *var* genes. Thus, upsB is associated with 22 out of 24 telomeric *var* genes, upsA is found with the two remaining telomeric *var* genes that are transcribed towards the telomere and with most telomere associated *var* genes (9 out of 12) which also point towards the telomere¹⁴¹. All 13 upsC sequences are associated with internal *var* clusters. Nearly all the telomeric *var* genes have an (A + T)-rich region approximately 2 kb upstream characterized by a number of poly(A) tracts as well as one or more copies of the consensus GGATCTAG. An analysis of the regions 1.0 kb downstream of *var* genes shows three sequence families, with members of one family being associated primarily with *var* genes

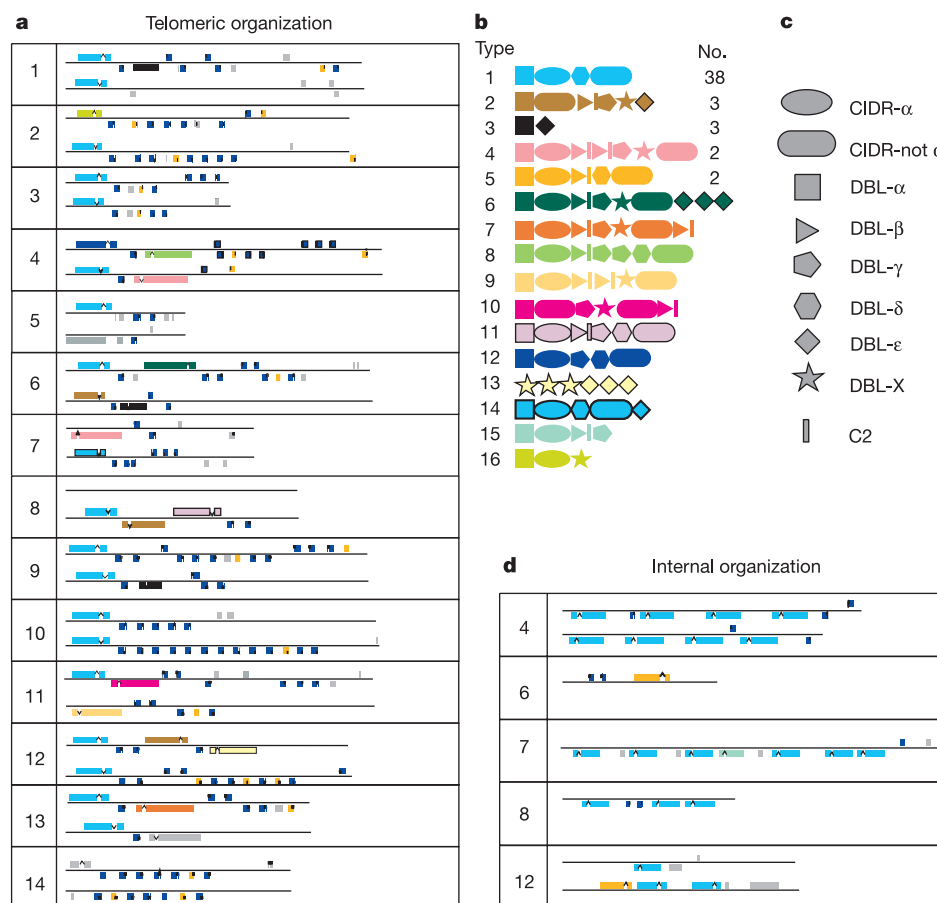


Figure 6 Organization of multi-gene families in *P. falciparum*. **a**, Telomeric regions of all chromosomes showing the relative positions of members of the multi-gene families: *rif* (blue) *stevor* (yellow) and *var* (colour coded as indicated; see **b** and **c**). Grey boxes represent pseudogenes or gene fragments of any of these families. The left telomere is shown above the right. Scale: ~ 0.6 mm = 1 kb. **b**, **c**, *var* gene domain structure. *var* genes contain three domain types: DBL, of which there are six sequence classes; CIDR, of

which there are two sequence classes; and conserved 2 (C2) domains (see text). The relative order of the domains in each gene is indicated (**c**). *var* genes with the same domain types in the same order have been colour coded as an identical class and given an arbitrary number for their type (**b**) and the total number of members of each class in the genome of *P. falciparum* clone 3D7. **d**, Internal multi-gene family clusters. Key as in **a**.

next to the telomeric repeats. The intron sequences within the *var* genes have been associated with locus specific silencing¹⁴². They vary in length from 170 to ~1,200 bp and are ~89% A/T. On the coding strand, at the 5' end the non-A/T bases are mainly G residues with 70% of sequences having the consensus TGT'TTGATATATA. The central regions are highly A-rich, and contain a number of semi-conserved motifs. The 3' region is comparably rich in C, with one or more copies in most genes of the sequence (TA)_n CCCATAAC-TACA. The 3' end has an extended and atypical splice consensus of ACANATATAGTTA(T)_n TAG. Sequences upstream of *rif* and *stevor* genes also have distinguishable upstream sequences, but a proportion of *rif* genes have the *stevor* type of 5' sequence. Because the majority of telomeric *var* genes share a similar structure and 5' and 3' sequences, they may form a unique group in terms of regulation of gene expression.

The most conserved *var* gene previously identified, which mediates adherence to chondroitin sulphate A in the placenta¹⁴³, is incomplete in 3D7 because of deletion of part of exon 1 and all of exon 2. This gene is located on the right telomere of chromosome 5 (Fig 6). The majority of *var* genes sequenced previously had been identified as they mediated adhesion to particular receptors, and most of them had more than four domains in exon 1. The fact that type 1 *var* genes containing only 4 domains predominate in the 3D7 genome suggests that previous analyses had been based on a highly biased sample. The significance of this in terms of the function of type 1 *var* genes remains to be determined.

Immune-evasion mechanisms such as clonal antigenic variation of parasite-derived red cell surface proteins (PfEMP1s, *rifins*) and modulation of dendritic cell function have been documented in *P. falciparum*^{131,132}. A putative homologue of human cytokine macrophage migration inhibitory factor (MIF) was identified in *P. falciparum*. In vertebrates, MIFs have been shown to function as immuno-modulators and as growth factors¹⁴⁴, and in the nematode *Brugia malayi*, recombinant MIF modulated macrophage migration and promoted parasite survival¹⁴⁵. An MIF-type protein in *P. falciparum* may contribute to the parasite's ability to modulate the immune response by molecular mimicry or participate in other host-parasite interactions.

Implications for vaccine development

An effective malaria vaccine must induce protective immune responses equivalent to, or better than, those provided by naturally acquired immunity or immunization with attenuated sporozoites¹⁴⁶. To date, about 30 *P. falciparum* antigens that were

identified via conventional techniques are being evaluated for use in vaccines, and several have been tested in clinical trials. Partial protection with one vaccine has recently been attained in a field setting¹⁴⁷. The present genome sequence will stimulate vaccine development by the identification of hundreds of potential antigens that could be scanned for desired properties such as surface expression or limited antigenic diversity. This could be combined with data on stage-specific expression obtained by microarray and proteomics^{14,15} analyses to identify potential antigens that are expressed in one or more stages of the life cycle. However, high-throughput immunological assays to identify novel candidate vaccine antigens that are the targets of protective humoral and cellular immune responses in humans need to be developed if the genome sequence is to have an impact on vaccine development. In addition, new methods for maximizing the magnitude, quality and longevity of protective immune responses will be required in order to produce effective malaria vaccines.

Concluding remarks

The *P. falciparum*, *Anopheles gambiae* and *Homo sapiens* genome sequences have been completed in the past two years, and represent new starting points in the centuries-long search for solutions to the malaria problem. For the first time, a wealth of information is available for all three organisms that comprise the life cycle of the malaria parasite, providing abundant opportunities for the study of each species and their complex interactions that result in disease. The rapid pace of improvements in sequencing technology and the declining costs of sequencing have made it possible to begin genome sequencing efforts for *Plasmodium vivax*, the second major human malaria parasite, several malaria parasites of animals, and for many related parasites such as *Theileria* and *Toxoplasma*. These will be extremely useful for comparative purposes. Last, this technology will enable sampling of parasite, vector and host genomes in the field, providing information to support the development, deployment and monitoring of malaria control methods.

In the short term, however, the genome sequences alone provide little relief to those suffering from malaria. The work reported here and elsewhere needs to be accompanied by larger efforts to develop new methods of control, including new drugs and vaccines, improved diagnostics and effective vector control techniques. Much remains to be done. Clearly, research and investments to develop and implement new control measures are needed desperately if the social and economic impacts of malaria are to be relieved. The increased attention given to malaria (and to other infectious diseases affecting tropical countries) at the highest levels of government, and the initiation of programmes such as the Global Fund to Fight AIDS, Tuberculosis and Malaria¹⁴⁸, the Multilateral Initiative on Malaria in Africa¹⁴⁹, the Medicines for Malaria Venture¹⁵⁰, and the Roll Back Malaria campaign¹⁵¹, provide some hope of progress in this area. It is our hope and expectation that researchers around the globe will use the information and biological insights provided by complete genome sequences to accelerate the search for solutions to diseases affecting the most vulnerable of the world's population. □

Methods

Sequencing, gap closure and annotation

The techniques used at each of the three participating centres for sequencing, closure and annotation are described in the accompanying Letters⁷⁻⁹. To ensure that each centres' annotation procedures produced roughly equivalent results, the Wellcome Trust Sanger Institute ('Sanger') and the Institute for Genomic Research ('TIGR') annotated the same 100-kb segment of chromosome 14. The number of genes predicted in this sequence by the two centres was 22 and 23; the discrepancy being due to the merging of two single genes by one centre. Of the 74 exons predicted by the two centres, 50 (68%) were identical, 9 (2%) overlapped, 6 (8%) overlapped and shared one boundary, and the remainder were predicted by one centre but not the other. Thus 88% of the exons predicted by the two centres in the 100-kb fragment were identical or overlapped.

Finished sequence data and annotation were transferred in XML (extensible markup language) format from Sanger and the Stanford Genome Technology Center to TIGR, and

Table 3 Domains of PfEMP1 proteins in *P. falciparum*

Domain type	Number of domains
DBL α	58
DBL β -C2	18
DBL γ	13
DBL δ	44
DBL ϵ	13
DBL-X	13
CIDR α	51
CIDR non- α	54
Preferred pairings	Frequency
DBL α -CIDR α	51/58
DBL β -C2	18/18
DBL δ -CIDR non- α	44/44
CIDR α -DBL δ	39/51
CIDR α -DBL β	10/51
DBL β -C2-DBL γ	10/18
DBL γ -DBL-X	8/13

Top, the total number of each DBL or CIDR domain type in intact *var* genes within the *P. falciparum* 3D7 genome. Bottom, the frequencies of the most common individual domain pairings found within intact *var* genes. The denominator refers to the total number of the first-named domains in intact *var* genes, and the numerator refers to the number of second-named domains found adjacent. See text for discussion of domain types.

made available to co-authors over the internet. Genes on finished chromosomes were assigned systematic names according to the scheme described previously⁵. Genes on unfinished chromosomes were given temporary identifiers.

Analysis of subtelomeric regions

Subtelomeric regions were analysed by the alignment of all of the chromosomes to each other using MUMmer²¹⁵² with a minimum exact match length ranging from 30 to 50 bp. Tandem repeats were identified by extracting a 90-kb region from the ends of all chromosomes and using Tandem Repeat Finder¹⁵³ with the following parameter settings: match = 2, mismatch = 7, indel = 7, pm = 75, pi = 10, minscore = 100, maxperiod = 500. Detailed pairwise alignments of internal telomeric blocks were computed with the ssearch program from the Fasta3 package¹⁵⁴.

Evolutionary analyses

Plasmodium falciparum proteins were searched against a database of proteins from all complete genomes as well as from a set of organelle, plasmid and viral genomes. Putative recently duplicated genes were identified as those encoding proteins with better BLASTP matches (based on E value with a 10⁻¹⁵ cutoff) to other proteins in *P. falciparum* than to proteins in any other species. Proteins of possible organellar descent were identified as those for which one of the top six prokaryotic matches (based on E value) was to either a protein encoded by an organelle genome or by a species related to the organelle ancestors (members of the *Rickettsia* subgroup of the α -Proteobacteria or cyanobacteria). Because BLAST matches are not an ideal method of inferring evolutionary history, phylogenetic analysis was conducted for all these proteins. For phylogenetic analysis, all homologues of each protein were identified by BLASTP searches of complete genomes and of a non-redundant protein database. Sequences were aligned using CLUSTALW, and phylogenetic trees were inferred using the neighbour-joining algorithms of CLUSTALW and PHYLIP. For comparative analysis of eukaryotes, the proteomes of all eukaryotes for which complete genomes are available (except the highly reduced *E. cuniculi*) were searched against each other. The proportion of proteins in each eukaryotic species that had a BLASTP match in each of the other eukaryotic species was determined, and used to infer a 'whole-genome tree' using the neighbour-joining algorithm. Possible eukaryotic conserved and specific proteins were identified as those with matches to all the complete eukaryotic genomes (10⁻³⁰ E-value cutoff) but without matches to any complete prokaryotic genome (10⁻¹⁵ cutoff).

Received 31 July; accepted 2 September 2002; doi:10.1038/nature01097.

1. Breman, J. G. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**, 1–11 (2001).
2. Greenwood, B. & Mutabingwa, T. Malaria in 2002. *Nature* **415**, 670–672 (2002).
3. Gallup, J. L. & Sachs, J. D. The economic burden of malaria. *Am. J. Trop. Med. Hyg.* **64**, 85–96 (2001).
4. Hoffman, S. L. *et al.* Funding for malaria genome sequencing. *Nature* **387**, 647 (1997).
5. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
6. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
7. Hall, N. *et al.* Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature* **419**, 527–531 (2002).
8. Gardner, M. J. *et al.* Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* **419**, 531–534 (2002).
9. Hyman, R. W. *et al.* Sequence of *Plasmodium falciparum* chromosome 12. *Nature* **419**, 534–537 (2002).
10. Foster, J. & Thompson, J. The *Plasmodium falciparum* genome project: a resource for researchers. *Parasitol. Today* **11**, 1–4 (1995).
11. Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
12. Su, X. Z. & Welles, T. E. Toward a high-resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* **33**, 430–444 (1996).
13. Lai, Z. *et al.* A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature Genet.* **23**, 309–313 (1999).
14. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
15. Lasender, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).
16. Watanabe, J., Sasaki, M., Suzuki, Y. & Sugano, S. FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, *Plasmodium falciparum*. *Nucleic Acids Res.* **29**, 70–71 (2001).
17. Gamain, B. *et al.* Increase in glutathione peroxidase activity in malaria parasite after selenium supplementation. *Free Radic. Biol. Med.* **21**, 559–565 (1996).
18. Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
19. Moriyama, E. N. & Powell, J. R. Codon usage bias and tRNA abundance in *Drosophila*. *J. Mol. Evol.* **45**, 514–523 (1997).
20. Duret, L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* **16**, 287–289 (2000).
21. Vaidya, A. B., Akella, R. & Suplick, K. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malaria parasite. *Mol. Biochem. Parasitol.* **35**, 97–107 (1989).
22. Vaidya, A. B., Lashgari, M. S., Polge, L. G. & Morrissey, J. Structural features of *Plasmodium* cytochrome *b* that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol. Biochem. Parasitol.* **58**, 33–42 (1993).
23. Tan, T. H., Pach, R., Crausaz, A., Ivens, A. & Schneider, A. tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. *Mol. Cell. Biol.* **22**, 3707–3717 (2002).

24. Tarassov, I. A. & Martin, R. P. Mechanisms of tRNA import into yeast mitochondria: an overview. *Biochimie* **78**, 502–510 (1996).
25. Wilson, R. J. M. *et al.* Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155–172 (1996).
26. Li, J., Wirtz, R. A., McConkey, G. A., Sattabongkot, J. & McCutchan, T. F. Transition of *Plasmodium vivax* ribosome types corresponds to sporozoite differentiation in the mosquito. *Mol. Biochem. Parasitol.* **65**, 283–289 (1994).
27. Waters, A. P. The ribosomal RNA genes of *Plasmodium*. *Adv. Parasitol.* **34**, 33–79 (1994).
28. Babiker, H. A., Creasey, A. M., Bayoumi, R. A., Walliker, D. & Arnot, D. E. Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 2. Drug resistance, molecular karyotypes and the *mdr1* genotype of recent isolates. *Trans. R. Soc. Trop. Med. Hyg.* **85**, 578–583 (1991).
29. Hinterberg, K., Mattei, D., Welles, T. E. & Scherf, A. Interchromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross. *EMBO J.* **13**, 4174–4180 (1994).
30. Hernandez, R. R., Hinterberg, K. & Scherf, A. Compartmentalization of genes coding for immunodominant antigens to fragile chromosome ends leads to dispersed subtelomeric gene families and rapid gene evolution in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **78**, 137–148 (1996).
31. Scherf, A. *et al.* Gene inactivation of Pf11-1 of *Plasmodium falciparum* by chromosome breakage and healing: identification of a gametocyte-specific protein with a potential role in gametocytogenesis. *EMBO J.* **11**, 2293–2301 (1992).
32. Day, K. P. *et al.* Genes necessary for expression of a virulence determinant and for transmission of *Plasmodium falciparum* are located on a 0.3-megabase region of chromosome 9. *Proc. Natl Acad. Sci. USA* **90**, 8292–8296 (1993).
33. Polge, L. G. & Ravetch, J. V. A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* **322**, 474–477 (1986).
34. Louis, E. J., Naumova, E. S., Lee, A., Naumov, G. & Haber, J. E. The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**, 789–802 (1994).
35. van Deutekom, J. C. *et al.* Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum. Mol. Genet.* **5**, 1997–2003 (1996).
36. Rudenko, G., McCulloch, R., Dirks-Mulder, A. & Borst, P. Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **80**, 65–75 (1996).
37. Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C. & Scherf, A. A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J.* **21**, 815–824 (2002).
38. Scherf, A., Figueiredo, L. M. & Freitas-Junior, L. H. *Plasmodium* telomeres: a pathogen's perspective. *Curr. Opin. Microbiol.* **4**, 409–414 (2001).
39. Vernick, K. D. & McCutchan, T. F. Sequence and structure of a *Plasmodium falciparum* telomere. *Mol. Biochem. Parasitol.* **28**, 85–94 (1988).
40. Oquendo, P. *et al.* Characterisation of a repetitive DNA sequence from the malaria parasite, *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **18**, 89–101 (1986).
41. De Bruin, D., Lanzer, M. & Ravetch, J. V. The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. *Proc. Natl Acad. Sci. USA* **91**, 619–623 (1994).
42. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
43. McFadden, G. I., Reith, M., Munhollan, J. & Lang-Unnasch, N. Plastid in human parasites. *Nature* **381**, 482–483 (1996).
44. Kohler, S. *et al.* A plastid of probable green algal origin in apicomplexan parasites. *Science* **275**, 1485–1489 (1997).
45. Fichera, M. E. & Roos, D. S. A plastid organelle as a drug target in apicomplexan parasites. *Nature* **390**, 407–409 (1997).
46. He, C. Y., Striepen, B., Pletcher, C. H., Murray, J. M. & Roos, D. S. Targeting and processing of nuclear-encoded apicoplast proteins in plastid segregation mutants of *Toxoplasma gondii*. *J. Biol. Chem.* **276**, 28436–28442 (2001).
47. Waller, R. F. *et al.* Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 12352–12357 (1998).
48. Surolia, N. & Surolia, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nature Med.* **7**, 167–173 (2001).
49. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
50. Sato, S. & Wilson, R. J. The genome of *Plasmodium falciparum* encodes an active delta-aminolevulinic acid dehydratase. *Curr. Genet.* **40**, 391–398 (2002).
51. Van Dooren, G. G., Su, V., D'Ombrian, M. C. & McFadden, G. I. Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J. Biol. Chem.* **277**, 23612–23619 (2002).
52. Wilson, R. J. Progress with parasite plastids. *J. Mol. Biol.* **319**, 257–274 (2002).
53. Stoebe, B. & Kowallik, K. V. Gene-cluster analysis in chloroplast genomics. *Trends Genet.* **15**, 344–347 (1999).
54. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18**, 418–426 (2001).
55. Roos, D. S. *et al.* Origin, targeting, and function of the apicomplexan plastid. *Curr. Opin. Microbiol.* **2**, 426–432 (1999).
56. Palmer, J. D. & Delwiche, C. F. Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl Acad. Sci. USA* **93**, 7432–7435 (1996).
57. Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* **19**, 1794–1802 (2000).
58. DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E. & Parsons, M. Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system. *J. Cell Sci.* **113** (Part 22), 3969–3977 (2000).
59. van Dooren, G. G., Schwartzbach, S. D., Osafune, T. & McFadden, G. I. Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophys. Acta* **1541**, 34–53 (2001).

60. Yung, S., Unnasch, T. R. & Lang-Unnasch, N. Analysis of apicoplast targeting and transit peptide processing in *Toxoplasma gondii* by deletional and insertional mutagenesis. *Mol. Biochem. Parasitol.* **118**, 11–21 (2001).
61. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I. & Schneider, G. Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* **280**, 19–26 (2001).
62. Vollmer, M., Thomsen, N., Wiek, S. & Seeber, F. Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast-localized, plant-type ferredoxin-NADP⁺ reductase and ferredoxin. *J. Biol. Chem.* **276**, 5483–5490 (2001).
63. Ralph, S. A., D’Ombrain, M. C. & McFadden, G. I. The apicoplast as an antimalarial drug target. *Drug Resist. Updat.* **4**, 145–151 (2001).
64. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977 (2000).
65. Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
66. Eisen, J. A. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* **8**, 163–167 (1998).
67. Adams, K. L., Daley, D. O., Whelan, J. & Palmer, J. D. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* **14**, 931–943 (2002).
68. Kanehisa, M., Goto, S., Kawashima, S. & Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30**, 42–46 (2002).
69. Sherman, I. W. in *Malaria Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 135–143 (ASM, Washington DC, 1998).
70. Buckwitz, D., Jacobasch, G., Gerth, C., Holzhtutter, H. G. & Thamm, R. A kinetic model of phosphofructokinase from *Plasmodium berghei*. Influence of ATP and fructose-6-phosphate. *Mol. Biochem. Parasitol.* **27**, 225–232 (1988).
71. Buckwitz, D., Jacobasch, G. & Gerth, C. Phosphofructokinase from *Plasmodium berghei*. Influence of Mg²⁺, ATP and Mg²⁺-complexed ATP. *Biochem. J.* **267**, 353–357 (1990).
72. Clarke, J. L., Scopes, D. A., Sodeinde, O. & Mason, P. J. Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. A novel bifunctional enzyme in malaria parasites. *Eur. J. Biochem.* **268**, 2013–2019 (2001).
73. Miclet, E. *et al.* NMR spectroscopic analysis of the first two steps of the pentose-phosphate pathway elucidates the role of 6-phosphogluconolactonase. *J. Biol. Chem.* **276**, 34840–34846 (2001).
74. Loyevsky, M. *et al.* An IRP-like protein from *Plasmodium falciparum* binds to a mammalian iron-responsive element. *Blood* **98**, 2555–2562 (2001).
75. Lang-Unnasch, N. Purification and properties of *Plasmodium falciparum* malate dehydrogenase. *Mol. Biochem. Parasitol.* **50**, 17–25 (1992).
76. Blum, J. J. & Ginsburg, H. Absence of α -ketoglutarate dehydrogenase activity and presence of CO₂-fixing activity in *Plasmodium falciparum* grown *in vitro* in human erythrocytes. *J. Protozool.* **31**, 167–169 (1984).
77. Fry, M. & Beesley, J. E. Mitochondria of mammalian *Plasmodium* spp. *Parasitology* **102**, 17–26 (1991).
78. Vaidya, A. B. in *Malaria: Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 355–368 (ASM, Washington DC, 1998).
79. Papa, S., Zanotti, F. & Gaballo, A. The structural and functional connection between the catalytic and proton translocating sectors of the mitochondrial F₁F₀-ATP synthase. *J. Bioenerg. Biomembr.* **32**, 401–411 (2000).
80. Sherman, I. W. in *Malaria: Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 177–184 (ASM, Washington DC, 1998).
81. de Macedo, C. S., Uhrig, M. L., Kimura, E. A. & Katzin, A. M. Characterization of the isoprenoid chain of coenzyme Q in *Plasmodium falciparum*. *FEMS Microbiol. Lett.* **207**, 13–20 (2002).
82. Trumpower, B. L. & Gennis, R. B. Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: the enzymology of coupling electron transfer reactions to transmembrane proton translocation. *Annu. Rev. Biochem.* **63**, 675–716 (1994).
83. Vaidya, A. B., McIntosh, M. T. & Srivastava, I. K. *Membrane Structure in Disease and Drug Therapy* (ed. Zimmer, G.) (Marcel Dekker, New York, 2000).
84. Perez-Martinez, X. *et al.* Subunit II of cytochrome c oxidase in Chlamydomonas algae is a heterodimer encoded by two independent nuclear genes. *J. Biol. Chem.* **276**, 11302–11309 (2001).
85. Murphy, A. D. & Lang-Unnasch, N. Alternative oxidase inhibitors potentiate the activity of atovaquone against *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **43**, 651–654 (1999).
86. Dieckmann, A. & Jung, A. Mechanisms of sulfadoxine resistance in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **19**, 143–147 (1986).
87. McConkey, G. A. Targeting the shikimate pathway in the malaria parasite *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **43**, 175–177 (1999).
88. Roberts, F. *et al.* Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **393**, 801–805 (1998).
89. Roberts, C. W. *et al.* The shikimate pathway and its branches in apicomplexan parasites. *J. Infect. Dis.* **185** (Suppl. 1), S25–S36 (2002).
90. Keeling, P. J. *et al.* Shikimate pathway in apicomplexan parasites. *Nature* **397**, 219–220 (1999).
91. Fitzpatrick, T. *et al.* Subcellular localization and characterization of chorismate synthase in the apicomplexan *Plasmodium falciparum*. *Mol. Microbiol.* **40**, 65–75 (2001).
92. Duncan, K., Edwards, R. M. & Coggins, J. R. The pentafunctional arom enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochem. J.* **246**, 375–386 (1987).
93. Rubin, H. *et al.* Cloning, sequence determination, and regulation of the ribonucleotide reductase subunits from *Plasmodium falciparum*: a target for antimalarial therapy. *Proc. Natl Acad. Sci. USA* **90**, 9280–9284 (1993).
94. Chakrabarti, D., Schuster, S. M. & Chakrabarti, R. Cloning and characterization of subunit genes of ribonucleotide reductase, a cell-cycle-regulated enzyme, from *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **90**, 12020–12024 (1993).
95. Krnjajski, Z., Gilberger, T. W., Walter, R. D. & Muller, S. The malaria parasite *Plasmodium falciparum* possesses a functional thioredoxin system. *Mol. Biochem. Parasitol.* **112**, 219–228 (2001).
96. Bonday, Z. Q., Dhanasekaran, S., Rangarajan, P. N. & Padmanaban, G. Import of host δ -aminolevulinic dehydratase into the malarial parasite: identification of a new drug target. *Nature Med.* **6**, 898–903 (2000).
97. Bonday, Z. Q., Taktani, S., Gupta, P. D. & Padmanaban, G. Heme biosynthesis by the malarial parasite. Import of δ -aminolevulinic dehydratase from the host red cell. *J. Biol. Chem.* **272**, 21839–21846 (1997).
98. Wilson, C. M., Smith, A. B. & Baylton, R. V. Characterization of the δ -aminolevulinic synthase gene homologue in *P. falciparum*. *Mol. Biochem. Parasitol.* **75**, 271–276 (1996).
99. Sato, S., Tews, I. & Wilson, R. J. Impact of a plastid-bearing endocytobiont on apicomplexan genomes. *Int. J. Parasitol.* **30**, 427–439 (2000).
100. Rohdich, F. *et al.* Biosynthesis of terpenoids. 2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) from *Plasmodium falciparum*. *Eur. J. Biochem.* **268**, 3190–3197 (2001).
101. Kemp, L. E., Bond, C. S. & Hunter, W. N. Structure of 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase: an essential enzyme for isoprenoid biosynthesis and target for antimicrobial drug development. *Proc. Natl Acad. Sci. USA* **99**, 6591–6596 (2002).
102. Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H. Jr Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301**, 75–100 (2000).
103. Woodrow, C. J., Burchmore, R. J. & Krishna, S. Hexose permeation pathways in *Plasmodium falciparum*-infected erythrocytes. *Proc. Natl Acad. Sci. USA* **97**, 9931–9936 (2000).
104. Hansen, M., Kun, J. F., Schultz, J. E. & Beitz, E. A single, bi-functional aquaglyceroporin in blood-stage *Plasmodium falciparum* malaria parasites. *J. Biol. Chem.* **277**, 4874–4882 (2002).
105. Elliott, J. L., Saliba, K. J. & Kirk, K. Transport of lactate and pyruvate in the intraerythrocytic malaria parasite, *Plasmodium falciparum*. *Biochem. J.* **355**, 733–739 (2001).
106. Rager, N., Mamoun, C. B., Carter, N. S., Goldberg, D. E. & Ullman, B. Localization of the *Plasmodium falciparum* PfNT1 nucleoside transporter to the parasite plasma membrane. *J. Biol. Chem.* **276**, 41095–41099 (2001).
107. Dyer, M., Wong, I. H., Jackson, M., Huynh, P. & Mikkelsen, R. Isolation and sequence analysis of a cDNA encoding an adenine nucleotide translocator from *Plasmodium falciparum*. *Biochim. Biophys. Acta* **1186**, 133–136 (1994).
108. McIntosh, M. T., Drowdzowicz, Y. M., Laroia, K., Rea, P. A. & Vaidya, A. B. Two classes of plant-like vacuolar-type H⁺-pyrophosphatases in malaria parasites. *Mol. Biochem. Parasitol.* **114**, 183–195 (2001).
109. Fidock, A. D. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6**, 861–871 (2000).
110. Desai, S. A., Bezrukov, S. M. & Zimmerberg, J. A voltage-dependent channel involved in nutrient uptake by red blood cells infected with the malaria parasite. *Nature* **406**, 1001–1005 (2000).
111. Eisen, J. A. & Hanawalt, P. C. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* **435**, 171–213 (1999).
112. Wood, R. D., Mitchell, M., Sgourou, J. & Lindahl, T. Human DNA repair genes. *Science* **291**, 1284–1289 (2001).
113. Haltiwanger, B. M. *et al.* DNA base excision repair in human malaria parasites is predominantly by a long-patch pathway. *Biochemistry* **39**, 763–772 (2000).
114. Critchlow, S. E. & Jackson, S. P. DNA end-joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394–398 (1998).
115. Freitas-Junior, L. H. *et al.* Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**, 1018–1022 (2000).
116. Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S. & Mitchell, G. H. A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. *Parasitol. Today* **16**, 427–433 (2000).
117. van Dooren, G. G., Waller, R. E., Joiner, K. A., Roos, D. S. & McFadden, G. I. Traffic jams: protein transport in *Plasmodium falciparum*. *Parasitol. Today* **16**, 421–427 (2000).
118. Wiser, M. E., Lanners, H. N., Bafford, R. A. & Favaloro, J. M. A novel alternate secretory pathway for the export of *Plasmodium* proteins into the host erythrocyte. *Proc. Natl Acad. Sci. USA* **94**, 9108–9113 (1997).
119. Albano, F. R. *et al.* A homologue of Sar1p localises to a novel trafficking pathway in malaria-infected erythrocytes. *Eur. J. Cell Biol.* **78**, 453–462 (1999).
120. Adisa, A., Albano, F. R., Reeder, J., Foley, M. & Tilley, L. Evidence for a role for a *Plasmodium falciparum* homologue of Sec31p in the export of proteins to the surface of malaria parasite-infected erythrocytes. *J. Cell Sci.* **114**, 3377–3386 (2001).
121. Hayashi, M. *et al.* A homologue of N-ethylmaleimide-sensitive factor in the malaria parasite *Plasmodium falciparum* is exported and localized in vesicular structures in the cytoplasm of infected erythrocytes in the brefeldin A-sensitive pathway. *J. Biol. Chem.* **276**, 15249–15255 (2001).
122. Knapp, B., Hundt, E. & Kupper, H. A. A new blood stage antigen of *Plasmodium falciparum* transported to the erythrocyte surface. *Mol. Biochem. Parasitol.* **37**, 47–56 (1989).
123. Sacher, M. *et al.* TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. *EMBO J.* **17**, 2494–2503 (1998).
124. Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *J. Exp. Med.* **159**, 1567–1575 (1984).
125. Weber, J. L. Interspersed repetitive DNA from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **29**, 117–124 (1988).
126. Su, Z. *et al.* The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100 (1995).
127. Baruch, D. I. *et al.* Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).
128. Smith, J. D. *et al.* Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
129. Cheng, Q. *et al.* *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97**, 161–176 (1998).
130. Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 9333–9338 (1999).
131. Kyes, S., Horrocks, P. & Newbold, C. I. Antigenic variation at the infected red cell surface in malaria. *Annu. Rev. Microbiol.* **55**, 673–707 (2001).
132. Urban, B. C. *et al.* *Plasmodium falciparum*-infected erythrocytes modulate the maturation of dendritic cells. *Nature* **400**, 73–77 (1999).
133. Pain, A. *et al.* Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc. Natl Acad. Sci. USA* **98**, 1805–1810 (2001).

134. Fried, M. & Duffy, P. E. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science* **272**, 1502–1504 (1996).
135. Udomsangpet, R. *et al.* *Plasmodium falciparum*-infected erythrocytes form spontaneous erythrocyte rosettes. *J. Exp. Med.* **169**, 1835–1840 (1989).
136. Bull, P. C. *et al.* Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Med.* **4**, 358–360 (1998).
137. Peterson, D. S., Miller, L. H. & Welles, T. E. Isolation of multiple sequences from the *Plasmodium falciparum* genome that encode conserved domains homologous to those in erythrocyte binding proteins. *Proc. Natl Acad. Sci. USA* **92**, 7100–7104 (1995).
138. Baruch, D. I. *et al.* Identification of a region of PfEMP1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36: conserved function with variant sequence. *Blood* **90**, 3766–3775 (1997).
139. Smith, J. D., Gamain, B., Baruch, D. I. & Kyes, S. Decoding the language of *var* genes and *Plasmodium falciparum* sequestration. *Trends Parasitol.* **17**, 538–545 (2001).
140. Smith, J. D. *et al.* Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proc. Natl Acad. Sci. USA* **97**, 1766–1771 (2000).
141. Voss, T. S. *et al.* Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum* *var* gene 5' flanking sequences. *Mol. Biochem. Parasitol.* **107**, 103–115 (2000).
142. Deitsch, K. W., Calderwood, M. S. & Welles, T. E. Malaria. Cooperative silencing elements in *var* genes. *Nature* **412**, 875–876 (2001).
143. Rowe, J. A., Kyes, S. A., Rogerson, S. J., Babiker, H. A. & Raza, A. Identification of a conserved *Plasmodium falciparum* *var* gene implicated in malaria in pregnancy. *J. Infect. Dis.* **185**, 1207–1211 (2002).
144. Lue, H., Kleemann, R., Calandra, T., Roger, T. & Bernhagen, J. Macrophage migration inhibitory factor (MIF): mechanisms of action and role in disease. *Microbes Infect.* **4**, 449–460 (2002).
145. Pastrana, D. V. *et al.* Filarial nematode parasites secrete a homologue of the human cytokine macrophage migration inhibitory factor. *Infect. Immun.* **66**, 5955–5963 (1998).
146. Richie, T. L. & Saul, A. Progress and challenges for malaria vaccines. *Nature* **415**, 694–701 (2002).
147. Bojang, K. A. *et al.* Efficacy of RTS,S/AS02 malaria vaccine against *Plasmodium falciparum* infection in semi-immune adult men in The Gambia: a randomised trial. *Lancet* **358**, 1927–1934 (2001).
148. Kapp, C. Global fund on AIDS, tuberculosis, and malaria holds first board meeting. *Lancet* **359**, 414 (2002).
149. Nchinda, T. C. Malaria: a reemerging disease in Africa. *Emerg. Infect. Dis.* **4**, 398–403 (1998).
150. Ridley, R. G. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* **415**, 686–693 (2002).
151. Nabarro, D. N. & Tayler, E. M. The “roll back malaria” campaign. *Science* **280**, 2067–2068 (1998).
152. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).
153. Benson, G. Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
154. Pearson, W. R. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* **132**, 185–219 (2000).
155. Glockner, G. *et al.* Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* **418**, 79–85 (2002).
156. Wood, V., Rutherford, K. M., Ivens, A., Rajandream, M. A. & Barrell, B. A re-annotation of the *Saccharomyces cerevisiae* genome. *Comp. Funct. Genom.* **2**, 143–154 (2001).
157. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
158. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
159. Scharfe, C. *et al.* MITOP, the mitochondrial proteome database: 2000 update. *Nucleic Acids Res.* **28**, 155–158 (2000).
160. Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
161. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
162. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
163. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
164. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
165. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium Yoelii yoelii*. *Nature* **419**, 512–519 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank our colleagues at The Wellcome Trust Sanger Institute, The Institute for Genomic Research, the Stanford Genome Technology Center, and the Naval Medical Research Center for their support. We thank J. Foster for providing markers for chromosome 14; R. Huestis and K. Fischer for providing RT–PCR data for chromosomes 2 and 3 before publication; A. Waters for assistance with ribosomal RNAs; S. Cawley for assistance with phat; and M. Crawford and R. Wang for discussions. This work was supported by the Wellcome Trust, the Burroughs Wellcome Fund, the National Institute for Allergy and Infectious Diseases, the Naval Medical Research Center, and the US Army Medical Research and Materiel Command.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.G. (e-mail: gardner@tigr.org). Sequences and annotation are available at the following websites: PlasmoDB (<http://plasmodb.org>), The Institute for Genomic Research (<http://www.tigr.org>), the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/Projects/Protozoa/>), and the Stanford Genome Technology Center (<http://www-sequence.stanford.edu/group/malaria>). Chromosome sequences were submitted to EMBL or GenBank with accession numbers AL844501–AL844509 (chromosomes 1, 3–9 and 13), AE001362.2 (chromosome 2), AE014185–AE014187 (chromosomes 10, 11 and 14) and AE014188 (chromosome 12).

Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*

Jane M. Carlton*, Samuel V. Angiuoli*, Bernard B. Suh*, Taco W. Kooij†, Mihaela Pertea*, Joana C. Silva*, Maria D. Ermolaeva*, Jonathan E. Allen*, Jeremy D. Selengut*, Hean L. Koo*, Jeremy D. Peterson*, Mihai Pop*, Daniel S. Kosack*, Martin F. Shumway*, Shelby L. Bidwell*, Shamira J. Shallom*, Susan E. van Aken*, Steven B. Riedmuller*, Tamara V. Feldblyum*, Jennifer K. Cho*, John Quackenbush*, Martha Sedegah§, Azadeh Shoaibi*, Leda M. Cummings*, Laurence Florens||, John R. Yates||, J. Dale Raine¶, Robert E. Sinden¶, Michael A. Harris#, Deirdre A. Cunningham*, Peter R. Preiser*, Lawrence W. Bergman**, Akhil B. Vaidya**, Leo H. van Lin†, Chris J. Janse†, Andrew P. Waters†, Hamilton O. Smith#, Owen R. White*, Steven L. Salzberg*, J. Craig Venter††, Claire M. Fraser*, Stephen L. Hoffman§, Malcolm J. Gardner* & Daniel J. Carucci§

* The Institute for Genomic Research, 9712 Medical Center Drive; and †† The Center for the Advancement of Genomics, 1901 Research Boulevard, Rockville, Maryland 20850, USA

† Department of Parasitology, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands

§ Naval Medical Research Center, Malaria Program (IDD), Silver Spring, Maryland 20910, USA

|| Department of Cell Biology, The Scripps Research Institute, La Jolla, California, 92037, USA

¶ Infection & Immunity Section, Department of Biological Sciences, Imperial College of Science, Technology & Medicine, London, SW7 2AZ, UK

Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850, USA

☆ Division of Parasitology, National Institute for Medical Research, London, UK

** Division of Molecular Parasitology, Department of Microbiology & Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania 19129, USA

Species of malaria parasite that infect rodents have long been used as models for malaria disease research. Here we report the whole-genome shotgun sequence of one species, *Plasmodium yoelii yoelii*, and comparative studies with the genome of the human malaria parasite *Plasmodium falciparum* clone 3D7. A synteny map of 2,212 *P. y. yoelii* contiguous DNA sequences (contigs) aligned to 14 *P. falciparum* chromosomes reveals marked conservation of gene synteny within the body of each chromosome. Of about 5,300 *P. falciparum* genes, more than 3,300 *P. y. yoelii* orthologues of predominantly metabolic function were identified. Over 800 copies of a variant antigen gene located in subtelomeric regions were found. This is the first genome sequence of a model eukaryotic parasite, and it provides insight into the use of such systems in the modelling of *Plasmodium* biology and disease.

For decades, the laboratory mouse has provided an alternative platform for infectious disease research where the pathogen under study is intractable to routine laboratory manipulation. Experimental study of the human malaria parasite *Plasmodium falciparum* is particularly problematic as the complete life cycle cannot be maintained *in vitro*. Four species of rodent malaria (*Plasmodium yoelii*, *Plasmodium berghei*, *Plasmodium chabaudi* and *Plasmodium vinckei*) isolated from wild thicket rats in Africa have been adapted to grow in laboratory rodents¹. These species reproduce many of the biological characteristics of the human malaria parasite. Many of the experimental procedures refined for use with *P. falciparum* were initially developed for rodent malaria species, a prime example being stable genetic transformation². Thus rodent models of malaria have been used widely and successfully to complement research on *P. falciparum*.

With the advent of the *P. falciparum* Genome Sequencing Project, undertaken by an international consortium of genome sequencing centres and malaria researchers, a series of initiatives has begun to generate substantial genome information from additional *Plasmodium* species³. We describe here the genome sequence of the rodent malaria parasite *P. y. yoelii* to fivefold genome coverage. We show that this partial genome sequencing approach, although limited in its application to the study of genome structure, has proved to be an effective means of gene discovery and of jump-starting experimental studies in a model *Plasmodium* species. Furthermore, we show

that despite the considerable divergence between the *P. y. yoelii* and *P. falciparum* genomes, sequencing and annotation of the former can substantially improve the accuracy and efficiency of annotation of the latter.

Plasmodium yoelii yoelii genome sequencing and annotation

We applied the whole-genome shotgun (WGS) sequencing approach, used successfully to sequence and assemble the first large eukaryotic genome⁴, to achieve fivefold sequence coverage of the genome of a clone of the 17XNL line of *P. y. yoelii* (Table 1). This level of coverage is expected to comprise 99% of the genome⁵ assuming random library representation. As with *P. falciparum*, the genomes of rodent malaria parasites are highly (A + T)-rich⁶, which adversely affects DNA stability in plasmid libraries. Consequently, all ~220,000 reads were produced from clones originating

Table 1 *Plasmodium yoelii yoelii* genome coverage statistics

Data	Component	Value
Genome	No. of contigs	5,687
	Mean contig size (kb)	3.6
	Max. contig size (kb)	51.5
	Cumulative contig length (Mb)	23.1
	No. of singletons	11,732
	No. of groups	2,906
	Max. group size (kb)	69.8
	Cumulative group size (Mb)	21.6
	No. of ESTs	13,080
	Average length (nucleotides)	497
Transcriptome	No. of gametocyte peptides	1,413
Proteome	No. of sporozoite peptides	677

‡ Present addresses: National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894, USA (L.M.C.); Genentech, San Francisco, California 94080, USA (J.K.C.); and Sanaria, 308 Argosy Drive, Gaithersburg, Maryland 20878, USA (S.L.H.).

from small (2–3 kilobases (kb)) insert libraries. Contigs were assembled using TIGR Assembler⁷. Contaminating mouse sequences, identified through similarity searches and found to comprise 10% of the total sequence data, were excluded from the analyses. Approximately three-quarters of the contigs could be placed into 2,906 'groups', each group consisting of two or more contigs known to be linked through paired reads as determined by Grouper software⁷. This produced an average group size of 7.4 kb, approximately 4 kb more than the average contig size. This group size is small compared with the group data produced by other partial eukaryotic genome projects, where extensive use of large insert (linking) libraries has enabled the construction of ordered and orientated 'scaffolds'⁸, and emphasizes the use of such linking libraries in partial genome projects. The genome size of *P. y. yoelii* is estimated to be 23 megabases (Mb), in agreement with karyotype data⁹.

Expression data from the *P. y. yoelii* transcriptome and proteome were generated to aid in gene identification and annotation of the contigs (Table 1). A total of 13,080 expressed sequence tag (EST) sequences generated from clones of an asexual blood-stage *P. y. yoelii* complementary DNA library¹⁰, in combination with other *P. yoelii* ESTs and transcript sequences available from public databases, were assembled and used to compile a gene index¹¹ of expressed *P. y. yoelii* sequences (<http://www.tigr.org/tdb/tgi/pygi/>). For protein expression data, multidimensional protein identification technology (MudPIT), which combines high-resolution liquid chromatography with tandem mass spectrometry and database searching, was applied to the gametocyte and salivary gland sporozoite proteomes of *P. y. yoelii*. A total of 1,413 gametocyte and 677 sporozoite peptides were recorded and used for the purposes of gene annotation.

We used two gene-finding programs, GlimmerMExon and Phat¹², to predict coding regions in *P. y. yoelii*. GlimmerMExon is based on the eukaryotic gene finder GlimmerM¹³, with modifications developed for analysing the short fragments of DNA that result from partial shotgun sequencing. Gene models based on GlimmerMExon and Phat predictions were refined using Combi-

ner. Annotation of predicted gene models used TIGR's fully automated Eukaryotic Genome Control suite of programs. Gene finding and subsequent annotation were limited to 2,960 contigs (each of which is over 2 kb in size), a subset of sequences that contains more than 20 Mb of the genome. A total of 5,878 complete genes and 1,952 partial genes (defined as genes lacking either an annotated start or stop codon) can be predicted from the nuclear genome data.

Comparative genome analysis

A comparison of several genome features of *P. falciparum* and *P. y. yoelii* is shown in Table 2, demonstrating that many similarities exist between the genomes. Besides the similarly extreme (G + C) compositions, both genomes contain a comparable number of predicted full-length genes, with the higher figure in *P. y. yoelii* due to an extremely high copy number of variant antigen genes (see below). Where differences between the genomes do exist, such as the (G + C) content of the coding portion of the genomes, incompleteness of the *P. y. yoelii* genome data, with the associated problems of accurate gene finding in both species, is likely to be a confounding factor. As an indication of this problem, analysis of *P. y. yoelii* proteomic data identified 83 regions of the genome apparently expressed during sporozoite and/or gametocyte stages but not assigned to a *P. y. yoelii* gene model (data not shown). Many of these peptide hits appear sufficiently close to a model as to indicate a fault with gene boundary prediction rather than a lack of gene prediction *per se*. However, as with the gene model prediction in *P. falciparum*, the gene models of *P. y. yoelii* should be considered preliminary and under revision.

Identifying orthologues of *P. falciparum* vaccine candidate proteins and proteins that are either targets of antimalarial drugs or involved in antimalarial drug resistance mechanisms is a primary goal of model malaria parasite genomics. Using BLASTP¹⁴ with a cutoff E value of 10^{-15} and no low-complexity filtering, 3,310 bi-directional orthologues (defined as genes related to each other through vertical evolutionary descent) can be identified in the full protein complement of *P. falciparum* (5,268 proteins) and the protein complement of *P. y. yoelii* translated from complete gene models (5,878 proteins). A list of vaccine candidate orthologues and orthologues of genes involved in antimalarial drug interactions identified from among the 3,310 orthologues and from additional BLAST analyses is shown in Table 3. Those genes that are not identifiable may either be absent from the partial genome data, or represent genes that have been lost or diverged sufficiently that they are undetectable through similarity searching.

Many of the candidate vaccine antigens under study in *P. falciparum* can be identified in *P. y. yoelii*, including orthologues of several asexual blood-stage antigens known to elicit immune responses in individuals exposed to natural infection (MSP1, AMA1, RAP1, RAP2). As immunity to *P. falciparum* blood-stage infection can be transferred by immune sera, identification of the targets of potentially protective antibody responses after natural infection can provide information beneficial to the selection of candidate antigens for malaria vaccines. We found several orthologues of known *P. falciparum* transmission-blocking candidates; in particular, members of the P48/45 gene family identified previously¹⁵ were confirmed.

We identified several *P. y. yoelii* orthologues of *P. falciparum* biochemical pathway components under study as targets for drug design (Table 3), most notably: (1) the 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DOXPR) gene whose product is inhibited by fosmidomycin in *P. falciparum* *in vitro* cultures and mice infected with *P. vinckeii*¹⁶; (2) enoyl-acyl carrier protein (ACP) reductase (FAB1) whose product is inhibited by triclosan in *P. falciparum* *in vitro* cultures and mice infected with *P. berghei*¹⁷; and (3) a gene encoding farnesyl transferase (FTASE), which is inhibited in cultures of *P. falciparum* treated with custom-designed peptidomimetics¹⁸. The rodent models of malaria have proved

Table 2 Comparison of genome features of *P. falciparum* and *P. y. yoelii*

Feature	<i>P. y. yoelii</i>	<i>P. falciparum</i>
Size (Mb)	23.1	22.9
No. of chromosomes	14	14
No. of gaps	5,812	93
Coverage*	5	14.5
(G + C) content (%)	22.6	19.4
No. of genes†	5,878	5,268
Mean gene length (bp)	1,298	2,283
Gene density (bp per gene)	2,566	4,338
Per cent coding	50.6	52.6
Genes with introns (%)	54.2	53.9
Genes with ESTs (%)	48.9	49.1
Gene products detected by proteomics (%)	18.2	51.8
Exons		
Mean no. per gene	2.0	2.4
(G + C) content (%)	24.8	23.7
Mean length (bp)	641	949
Introns		
(G + C) content (%)	21.1	13.5
Mean length (bp)	209	179
Total length (bp)	1,687,689	1,323,509
Intergenic regions		
(G + C) content (%)	20.7	13.6
Mean length (bp)	859	1,694
RNAs		
No. of tRNA genes‡	39	43
No. of 5S rRNA genes	3	3
No. of 5.8S, 18S and 28S rRNA units	4	7
Mitochondrial genome		
(G + C) content (%)	31	31
Apicoplast genome		
(G + C) content (%)	15	14

*Average number of sequence reads per nucleotide.

†Total number of full-length genes.

‡The smaller number reflect the partial nature of the *P. y. yoelii* genome data.

Table 3 *P. y. yoelii* orthologues of *P. falciparum* candidate vaccine and drug interaction genes

<i>P. falciparum</i> gene	<i>Pf</i> chromosome	ST location*	<i>Pf</i> locus	<i>Py</i> locus
Candidate vaccine antigens				
Ring-infected erythrocytic surface antigen 1, <i>resa1</i>	1	Yes	PFA0110w	Not identified
Merozoite surface protein 4, <i>msp4</i>	2	No	PFB0310c	PY07543†
Merozoite surface protein 5, <i>msp5</i>	2	No	PFB0305c	PY07543†
Liver stage antigen 3, <i>lsa3</i>	2	No	PFB0915w	Not identified
Merozoite surface protein 2, <i>lsa3</i>	2	No	PFB0300c	Not identified
Transmission-blocking target antigen 230, <i>Pfs230</i>	2	No	PFB0405w	PY03856
Circumsporozoite protein, <i>csp</i>	3	No	MAL3P2.11	PY03168
Rhoptry-associated protein 2, <i>rap2</i>	5	Yes	PFE0080c	PY03918
Sporozoite surface antigen, <i>starp</i>	7	Yes	PF07_0006	Not identified
Merozoite surface protein 1, <i>msp1</i>	9	No	PF1475w	PY05748
Liver stage antigen 1, <i>lsa1</i>	10	No	PF10_0356	Not identified
Merozoite surface protein 3, <i>msp3</i>	10	No	PF10_0345	Not identified
Glutamate-rich protein, <i>glurp</i>	10	No	PF10_0344	Not identified
Ookinete surface protein 25, <i>Pfs25</i>	10	No	PF10_0303	PY00523
Ookinete surface protein 28, <i>Pfs28</i>	10	No	PF10_0302	PY00522
Erythrocyte membrane-associated 332 antigen, <i>Pf332</i>	11	No	PF11_0507	PY06496
Apical membrane antigen 1, <i>ama1</i>	11	No	PF11_0344	PY01581
Exported protein 1, <i>exp1</i>	11	No	PF11_0224	Not identified
Surface sporozoite protein 2, <i>ssp2</i>	13	No	PF13_0201	PY03052
Sexual-stage-specific surface antigen 48/45, <i>Pfs48/45</i>	13	No	PF13_0247	PY04207
Rhoptry-associated protein 1, <i>rap1</i>	14	Yes	PF14_0637	PY00622
Candidate drug interaction genes				
Dihydrofolate reductase, <i>dhfr</i>	4	No	PFD0830w	PY04370
Multidrug resistance protein 1, <i>pfmdr1</i>	5	No	PFE1150w	PY00245
Translationally controlled tumour protein, <i>tctp</i>	5	No	PFE0545c	PY04896
Farnesyl transferase, <i>ftase</i>	5	No	PFE0970w	PY06214
Enoyl-acyl carrier reductase, <i>fabi</i>	6	No	MAL6P1.275	PY03846
Dihydro-protate dehydrogenase, <i>dhod</i>	6	No	MAL6P1.36	PY02580
Chloroquine-resistance transporter, <i>pfcr1</i>	7	No	MAL7P1.27	PY05061
Dihydropteroate synthase, <i>dhps</i>	8	No	PF08_0095	PY02226
Lactate dehydrogenase, <i>ldh</i>	13	No	PF13_0141	PY03885
DOXP reductoisomerase, <i>doxpr</i>	14	No	PF14_0641	PY05578

A full listing of all orthologues can be found as Table A in the Supplementary Information. *Pf*, *P. falciparum*; *Py*, *P. y. yoelii*.
*ST, subtelomeric. Defined as >75% of the distance from the centre to the end of the *P. falciparum* chromosome.
†Homologue of *P. falciparum* *msp4* and *msp5* genes found as a single gene *msp4/5* in *P. y. yoelii* and other rodent malaria species⁶².

invaluable both for the study of potency of new antimalarial compounds *in vivo*, and for the elucidation of mechanisms of antimalarial drug resistance.

We applied the Gene Ontology (GO) gene classification system¹⁹, which uses a controlled vocabulary to describe genes and their function, to indicate which classes of gene among the 3,310 orthologues might differ in number between *P. falciparum* and *P. y. yoelii* (Fig. 1). A similar proportion of proteins were identified for most of the GO classes between the two species, with the caveat that fewer total numbers of proteins were identified in *P. y. yoelii* owing to the partial nature of the genome data for this species. However, proteins allocated to the physiological processes, cell invasion and adhesion, and cell communication categories were significantly reduced in *P. y. yoelii*. These classes contain members of three multigene families whose genes are found predominantly in the subtelomeric regions of *P. falciparum* chromosomes: PfEMP1, the protein product of the *var* gene family known to be involved in antigenic variation, cyto-adherence and rosetting, and rifins and stevors, which are clonally variant proteins possibly involved in antigenic variation and evasion of immune responses (reviewed in ref. 20). Apparently, *P. falciparum* has generated species-specific, subtelomeric genes involved in host cell invasion, adhesion and antigenic variation, homologues of which are not found in the *P. y. yoelii* genome.

Gene families of unique interest in the *P. y. yoelii* genome

The largest family of genes identified in the *P. y. yoelii* genome is the *yir* gene family, homologues of the *vir* multigene family recently described in the human malaria parasite *Plasmodium vivax*²¹ and in other species of rodent malaria²². In *P. vivax*, an estimated 600–1,000 copies of the subtelomerically located *vir* gene encode proteins that are immunovariant in natural infections, indicating a possible functional role in antigenic variation and immune evasion. Within the *P. y. yoelii* genome data, 838 *yir* genes (693

full genes and 145 partial genes) are present (Table 4; see also Supplementary Figs A and B). Almost 75% of the annotated contigs identified as containing subtelomeric sequences (see below) contain *yir* genes, many arranged in a head-to-tail fashion. Expression data indicate that *yir* genes are expressed during sporozoite, gametocyte and erythrocytic stages of the parasite, similar to the expression pattern seen with *P. falciparum* *var* and *rif* genes²³. Preliminary

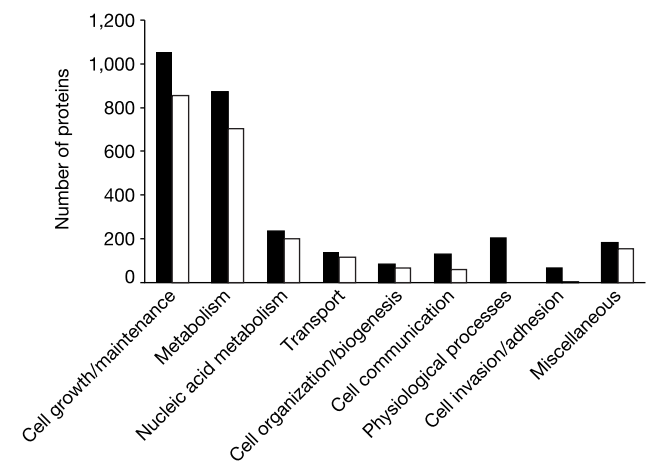


Figure 1 Functional classification comparison between *P. falciparum* and *P. y. yoelii* proteins. We compared the GO terms of proteins assigned to ‘biological process’ for the orthologous genes identified between the two species. The process group contains 3,041 *P. falciparum* annotations (filled bars), and 2,161 reciprocal annotations are shown for *P. y. yoelii* (open bars). Ten GO classes with similar numbers of *P. falciparum* and *P. y. yoelii* proteins in each are assigned as ‘miscellaneous’; that is, cell cycle, external stimulus response, stress response, signal transduction, homeostasis, developmental processes, cell proliferation, membrane fusion, death, cell motility.

Table 4 Paralogous gene families in *P. y. yoelii*

Gene family	No.	Name	HMM ID	Location in <i>Py</i>	<i>Py</i> expression*	<i>Pf</i> locus	TM/SP†
<i>yir/bir/cir</i>	838	Variant antigen family	TIGR01590	Subtelomeric	Gmt, spz, bs	None	P/A
<i>235 kDa</i>	14	Reticulocyte binding family	TIGR01612	Subtelomeric	Gmt, spz, bs	PFD0110w, MAL13P1.176, PF13_0198, PFL2520w, PFD0110w	P/A
<i>pyst-a</i>	168	Hypothetical	TIGR01599	Subtelomeric	Gmt, spz	PF14_0604	A/A
<i>pyst-b</i>	57	Hypothetical	TIGR01597	Subtelomeric	Bs	None	P/A
<i>pyst-c</i>	21	Hypothetical	TIGR01601, TIGR01604	Subtelomeric	Bs	None	P/P
<i>pyst-d</i>	17	Hypothetical	TIGR01605	Subtelomeric	Gmt	None	P/P
<i>etramp</i>	11	Early transcribed membrane protein family	TIGR01495	Subtelomeric	Gmt, spz, bs	PF13_0012, PF14_0016, PF11_0040, PFB0120w, PF10_0323, MAL12P1.387, PF11_0039, PFL1095c, PF10_0019, PF1745c, PFE1590w, PF10_0164, MAL8P1.6, PFA0195w, PFL0065w, PF14_0729, PFL2530w, PF10_0379, PF14_0738, PF14_0017, PF14_0737, PFI800w, PFI1775w, PF07_0040, PF07_0005, PFA0120c, PFC0110w, PFC0120w, PFI1730w, PFI1710w, PFB0935w	P/P
<i>pst-a</i>	12	Hydrolase family	TIGR01607	Subtelomeric	Gmt, spz		A/A
<i>rhoph1/clag</i>	2	Rhoptry H1/ cyto-adherence-linked asexual gene family	PF03805	Subtelomeric	Gmt, bs		A/P

*Found in, but not limited to: gmt, gametocyte life stage; spz, sporozoite life stage; bs, asexual blood stage.

†TM, transmembrane domain; SP, signal peptide; P, predicted; A, absent. TM and SP predictions were identical for *P. falciparum* and *P. y. yoelii* members of the same gene family. (See ref. 30 for details regarding TM and SP prediction algorithms.)

results using antibodies developed against the conserved regions of the protein have confirmed protein localization at the surface of the infected red blood cell (D.A.C. *et al.*, manuscript in preparation). The number of gene copies in the *P. y. yoelii* genome, the localization and stage-specific expression of gene members, as well as the existence of homologues in other *Plasmodium* species, make this gene family a prime target for the study of mechanisms of immune evasion.

A maximum of 14 members of the *Py235* multigene family can be identified among the *P. y. yoelii* protein data (Table 4). This family expresses proteins that localize to rhoptries (organelles that contain proteins involved in parasite recognition and invasion of host red blood cells). *Py235* genes exhibit a newly discovered form of clonal antigenic variation, whereby each individual merozoite derived from a single parent schizont has the propensity to express a different *Py235* protein²⁴. Closely related homologues of the *Py235* gene family have been found in other rodent malaria species, and more distantly related homologues have been found in *P. vivax*²⁵ and *P. falciparum*²⁶. The gene copy number identified in the current data set is less than has been predicted in other *P. y. yoelii* lines (30–50 per genome). This could reflect real differences in copy number between lines, but more probably suggests an error in the original estimate or misassembly of extremely closely related sequences. Almost all of the *Py235* genes are found on contigs identified as subtelomeric in the *P. y. yoelii* genome (see Supplementary Fig. C).

Four further paralogous gene families, *pyst-a* to *-d*, are specific to *P. y. yoelii* (Table 4). The *pyst-a* family deserves mention, as it is homologous to a *P. chabaudi* glutamate-rich protein²⁷ and to a single hypothetical gene on *P. falciparum* chromosome 14, suggesting expansion of this family in the rodent malaria species from a common ancestral *Plasmodium* gene. Two paralogous gene families containing multiple members are homologous to multi-gene families identified in *P. falciparum*. Gene members of one family, *etramp* (early transcribed membrane protein), have previously been identified in *P. falciparum*²⁸ and in *P. chabaudi* where a single member has been identified and localized to the parasitophorous vacuole membrane²⁹.

Telomeres and chromosomal exchange in subtelomeric regions

The telomeric repeat in *P. y. yoelii* is AACCCCTG, which differs from

the *P. falciparum* telomeric repeat AACCCCTA by one nucleotide. A total of 71 contigs were found to contain telomeric repeat sequences arranged in tandem, with the largest array consisting of 186 copies. The *P. y. yoelii* subtelomeric chromosomal regions show little repeat structure compared with those of *P. falciparum*. A survey of tandem repeats in the entire genome found only a few in the telomeric or subtelomeric regions, specifically a 15 base pair (bp) (45 copies) and a 31-bp (up to 10 copies), both of which were found on multiple contigs, and a 36-bp repeat that occurred on one contig. No repeat element that corresponds to Rep20, a highly variable 21-bp unit that spans up to 22 kb in *P. falciparum* telomeres, was found.

The telomeric and subtelomeric regions of *P. y. yoelii* contigs show extensive large-scale similarity, indicating that these regions undergo chromosomal exchange similar to that reported for *P. falciparum* (see ref. 30). The longest subtelomeric contig is approximately 27 kb (see Supplementary Fig. C) and is homologous to other subtelomeric contigs across its entire length, indicating that the region of chromosomal exchange extends at least this distance into the subtelomeres. Recent data have shown that clustering of telomeres at the nuclear periphery in asexual and sexual stage *P. falciparum* parasites may promote sequence exchange between members of subtelomeric virulence genes on heterologous chromosomes, resulting in diversification of antigenic and adhesive phenotypes (see ref. 31 for review). The suggestion of extensive chromosome exchange in *P. y. yoelii* indicates that a similar system for generating antigenic diversity of the *yir*, *Py235* and other gene families located within subtelomeric regions may exist.

A genome-wide synteny map

The *Plasmodium* lineage is estimated to have arisen some 100–180 million years ago³², and species of the parasite are known to infect birds, mammals and reptiles³³. On the basis of the analysis of small subunit (SSU) ribosomal RNA sequences, the closest relative to *P. falciparum* is *Plasmodium reichenowi*, a parasite of chimpanzees, with the rodent malaria species forming a distinct clade^{34,35}. Early gene mapping studies have shown that regions of gene synteny exist between species of rodent malaria⁹ and between human malaria species^{36,37}, despite extensive chromosome size polymorphisms between homologous chromosomes³⁸. This level of gene synteny seems to decrease as the phylogenetic distance between *Plasmodium* species increases³⁹. Before the *Plasmodium* genome sequencing

projects, the degree to which conservation of synteny extended across *Plasmodium* genomes was not fully apparent.

Using the *P. falciparum* and *P. y. yoelii* genome data, we have constructed a genome-wide syntenic map between the species. To avoid confounding factors inherent in DNA-based analyses of (A + T)-rich genomes, we first calculated the protein similarity between all possible protein-coding regions in both data sets using MUMmer⁴⁰. Sensitivity was ensured through the use of a minimum word match length of five amino acids chosen to identify seed maximal unique matches (MUMs). By comparison, the recent human–mouse synteny analysis used a match length of 11 (ref. 8). Using this method, which is independent of gene prediction data, 2,212 sequences could be aligned (tiled) to *P. falciparum* chromosomes, representing a cumulative length of 16.4 Mb of sequence, or over 70% of the *P. y. yoelii* genome (see Supplementary Table C). The per cent of each *P. falciparum* chromosome covered with *P. y. yoelii* matches varies from 12% (chromosome 4) to 22% (chromosomes 1 and 14), with an average of about 18%. The spatial arrangement of the tiling paths (see Fig. 1 in ref. 30) confirms previous suggestions⁹ that most of the conserved matches are found within the body of *Plasmodium* chromosomes, and confirms the absence of *var*, *rif* and *stevor* homologues in the *P. y. yoelii* genome.

Although the tiling paths indicate the degree of conservation of gene order between *P. falciparum* and *P. y. yoelii*, longer stretches of contiguous *P. y. yoelii* sequence are necessary to examine this feature in depth. Accordingly, we carried out linkage of many *P. y. yoelii* assemblies adjacent to each other along the tiling paths. First, 1,050 adjacent contigs were linked on the basis of paired reads as determined by Grouper software. Second, *P. y. yoelii* ESTs were aligned to the tiling paths, and those found to overlap sequences adjacent in the tiling path were used as evidence to link a further 236 *P. y. yoelii* sequences. Third, amplification of the sequence between adjacent contigs in the tiling paths linked a further 817 assemblies. Linkage of *P. y. yoelii* sequences by these methods resulted in the formation of 457 syntenic groups from 2,212 original contigs, ranging in length from a few kilobases to more than 800 kb. Syntenic groups were assigned to a *P. y. yoelii* chromosome where possible through the use of a partial physical map⁹. Thus, long contiguous sections of the *P. y. yoelii* genome with accompanying *P. y. yoelii* chromosomal location can be assigned to each *P. falciparum* chromosome (see Fig. 1 in ref. 30). The degree of conservation of gene order between the species was examined using ordered and orientated syntenic groups and Position Effect software. Of 4,300 *P. y. yoelii* genes within the syntenic groups, 3,145 (73%) were found to match a region of *P. falciparum* in conserved order.

One section of the syntenic map between *P. falciparum* and *P. y.*

yoelii in particular—associated with *P. falciparum* chromosomes 4 and 10 and *P. y. yoelii* chromosome 5—provides a detailed snapshot of synteny between the species. Chromosome 5 of *P. y. yoelii* has received particular attention owing to the localization of a number of sexual-stage-specific genes to it⁴¹, and because truncated versions of the chromosome are found in lines of the rodent malaria parasite *P. berghei*, which is defective in gametocytogenesis⁴². Genomic resources available for *P. berghei* chromosome 5 include chromosome markers and long-range restriction maps⁴¹. Exploiting the high level of synteny of rodent malaria parasite chromosomes⁹, these tools were applied in combination with further mapping studies to close the syntenic map of chromosome 5 of *P. y. yoelii* (Fig. 2).

Approximately 0.8 Mb of *P. y. yoelii* chromosome 5 (estimated total length of 1.5 Mb) could be linked into one group that is syntenic to *P. falciparum* chromosome 10 and *P. falciparum* chromosome 4. From a total of 243 genes predicted in the syntenic region of *P. falciparum* chromosome 10, and 34 genes predicted in the syntenic region of chromosome 4, 171 (70%) and 22 (65%) of these, respectively, have homologues along *P. y. yoelii* chromosome 5 that appear in the same order. Pairs of homologous genes that map to regions of conserved synteny between *P. y. yoelii* and *P. falciparum* are probably orthologues, confirmed by the finding that most of these homologous pairs are also reciprocal best matches between the *P. falciparum* and *P. y. yoelii* proteins. Genes in the synteny gap on chromosome 10 (Fig. 2) include a glutamate-rich protein, S antigen, MSP3, MSP6 and liver stage antigen 1, several of which are prime vaccine antigen candidates in *P. falciparum*. Genes in the synteny gap on chromosome 4 include four *var* and two *rif* genes, which make up one of the four internal clusters of *var/rif* genes found in *P. falciparum* (see ref. 30). A series of uncharacterized hypothetical genes occur on the contigs that overlap these regions in *P. y. yoelii*.

An intriguing finding from the study of chromosome 5 has been the analysis of the syntenic break point between *P. falciparum* chromosomes 4 and 10. The final *P. y. yoelii* contig in the tiling path with significant synteny to *P. falciparum* chromosome 10 also contains the external transcribed sequence (ETS) of the SSU rRNA C unit. The synteny resumes on *P. falciparum* chromosome 4 in a *P. y. yoelii* contig that also contains the ETS of the large subunit (LSU) of the same rRNA unit. (No rRNA unit sequences are located on *P. falciparum* chromosomes 4 and 10; matches to contigs containing these genes occur in coding regions of other genes.) Both *P. y. yoelii* contigs are linked to each other through a third contig that contains the remaining elements (SSU, 5.8S, LSU, and internal transcribed sequences 1 and 2) of the complete rRNA unit (Fig. 2). Thus it seems that the break in synteny between *Plasmo-*

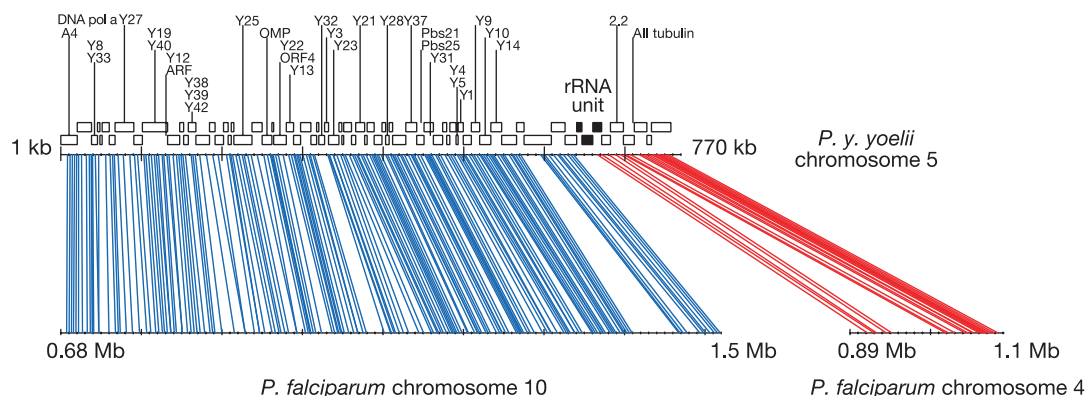


Figure 2 Conservation of gene synteny between *P. y. yoelii* chromosome 5 and *P. falciparum* chromosomes 4 and 10. Physical marker data used to confirm contig order in the tiling path of *P. y. yoelii* chromosome 5 are shown above the contigs (open boxes).

Each coloured line represents a pair of orthologous genes present in the two species shown anchored to its respective location in the two genomes. Contigs containing the *P. y. yoelii* rRNA unit are shown as filled boxes.

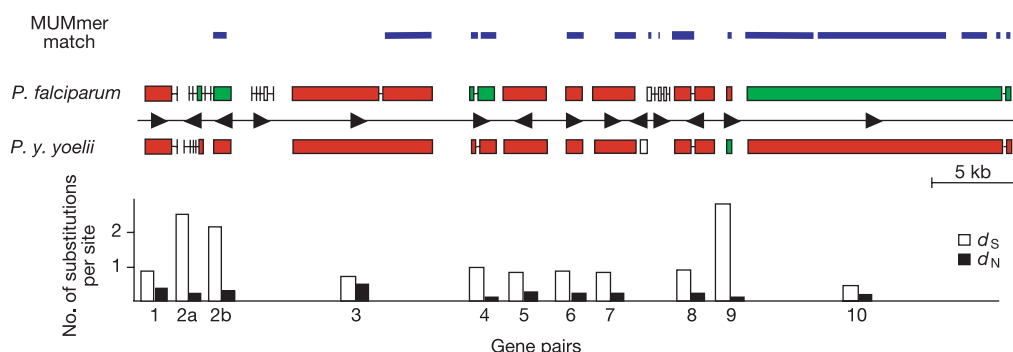


Figure 3 Global alignment scheme of a syntenic region between *P. falciparum* and *P. y. yoelii* encompassing ten orthologous gene pairs and nine intergenic regions. White boxes represent genes that have no orthologue and were excluded from analysis; green boxes represent gene models that were refined; red boxes represent unaltered gene models; arrowheads represent gene orientation on the DNA molecule. Clusters of

MUMmer matches between the two species are represented as thick blue lines. For the ten orthologous gene pairs, synonymous mutations per synonymous site (d_S , open bars) and non-synonymous mutations per non-synonymous site (d_N , filled bars) were estimated and plotted.

dium chromosomes has occurred within a single rRNA unit, a phenomenon first reported in prokaryotes⁴³. Six rRNA units reside as individual operons on *P. falciparum* chromosomes 1, 5, 7, 8, 11 and 13 respectively (ref. 30), in contrast to rodent malaria species that have four⁴⁴. Intriguingly, breaks in the synteny between *P. y. yoelii* and *P. falciparum* can be mapped to almost all rRNA unit loci on the *P. falciparum* chromosomes (see Fig. 1 of ref. 30). A full analysis of this potential phenomenon is outside the scope of this study, but these results provide preliminary evidence for one possible mechanism underlying synteny breakage that may have occurred during evolution of the *Plasmodium* genus—that of chromosome breakage and recombination at sites of rRNA units.

Comparative alignment of syntenic regions

Recent comparative studies have revealed that the fine detail of short stretches of the rodent and human malaria parasite genomes is remarkably conserved⁴⁵, and that such comparisons are useful for gene prediction and evolutionary studies. Accordingly, we used a comparison of the longest assembly of *P. y. yoelii* (MALPY00395, 51.3 kb) and its syntenic region in *P. falciparum* (chromosome 7, at coordinates 1,131–1,183 kb) as a case study for a preliminary evolutionary analysis of the two genomes. Gene prediction programs run against these two regions identified 11 genes in the syntenic region of both species (Fig. 3), eight of which are orthologous gene pairs (genes 1, 3–8 and 10). The structures of two additional gene pairs (genes 2a/b and 9) were refined through manual curation of erroneous gene boundaries. Three hypothetical genes, two in *P. falciparum* and one in *P. y. yoelii*, had no discernible orthologue in the other species; the presence of multiple stop codons in these areas suggests that the genes may have become pseudogenes. A global alignment at the DNA level of the syntenic region (Fig. 3) reveals the similarity between species in intergenic regions to be almost negligible, as mirrored in similar syntenic comparisons of mouse and human^{46,47}. Moreover, the mutation saturation observed in intergenic regions suggests that ‘phylogenetic footprinting’ can be used to identify conserved motifs between species that may be involved in gene regulation.

In contrast to intergenic regions, the similarity between species in coding regions is relatively high. The average number of non-synonymous substitutions per non-synonymous site, d_N , between the two species is 26% ($\pm 12\%$). Synonymous sites, d_S , are saturated (average $d_S > 1$), which supports the lack of similarity observed within intergenic regions. These values are considerably higher than those reported for human–rodent comparisons, which are approximately 7.5% and 45% for non-synonymous and synonymous substitutions, respectively⁴⁸. The cause of such apparent disparities

remains unknown, but may be a consequence of extreme genome composition or the short generation time of the parasite.

Rodent malaria species as models for *P. falciparum* biology

The usefulness of rodent malaria species as models for the study of *P. falciparum* is controversial. It is apparent that rodent models are the first port of call when preliminary *in vivo* evidence of antimalarial drug efficacy, immune response to vaccine candidates, and life-cycle adaptations in the face of drug or vaccine challenge are required. Different species of malaria parasite have developed different mechanisms of resistance to the antimalarial drug chloroquine, despite a similar mode of action of the drug (reviewed in ref. 49). It seems that mechanisms developed by the parasite to evade an inhospitable environment, whether caused by antimalarial drugs or the host immune system, may differ widely from species to species. A model involving evolution of different genes in *Plasmodium* species as a response to different host environments is consistent with the comparison of the *P. falciparum* and *P. y. yoelii* genomes presented here; conservation of synteny between the two species is high in regions of housekeeping genes, but not in regions where genes involved in antigenic variation and evasion of the host immune system are located. On the one hand, this can be interpreted as a blow to the systematic identification of all orthologues of antigen genes between *P. falciparum* and *P. y. yoelii* that could be used in the design of a malaria vaccine. On the other hand, a picture is emerging of selecting a model malaria species based on the complement of genes that best fit the phenotypic trait under study. Thus the presence of homologues of the *yir* family may make *P. y. yoelii* an attractive model for studying antigenic variation in *P. vivax*. Furthermore, identification of orthologues in the genomes of relatively distant rodent and human malaria parasites will facilitate finding orthologues in other model malaria species, for example monkey models of malaria such as *Plasmodium knowlesi*. □

Methods

Genome and EST sequencing

Plasmodium yoelii yoelii 17XNL line⁵⁰, selected from an isolate taken from the blood of a wild-caught thick rat in the Central African Republic⁵¹, is a non-lethal strain with a preference for development in reticulocytes. Clone 1.1 was obtained through serial dilution of sporozoites. Parasites were grown in laboratory mice no more than three blood passages from mosquito passage to limit chromosome instability, collected by exsanguination into heparin, and host mouse leukocytes were removed by filtration. Small insert libraries (average insert size 1.6 kb) were constructed in pUC-derived vectors after nebulization of genomic DNA. DNA sequencing of plasmid ends used ABI Big Dye terminator chemistry on ABI3700 sequencing machines. A total of 222,716 sequences (82% success rate), averaging 662 nucleotides in length, were assembled using TIGR Assembler⁷. BLASTN of the *P. y. yoelii* contigs and singletons against the complete set of

Celera mouse contigs⁸, using a cutoff of 90% identity over 100 nucleotides, identified contaminating mouse sequences that were subsequently removed. Contigs were assigned to groups using Grouper⁵². Each contig was assigned an identifier in the format 'MALPY00001'.

Proteomic analysis

MudPIT technology and methods were as described in ref. 23. Sporozoites of *P. y. yoelii* were dissected from infected *Anopheles stephensi* mosquito salivary glands, and *P. y. yoelii* gametocytes were prepared as described⁵³. Cellular debris from uninfected mosquitoes and mouse erythrocytes were analysed as controls. Tandem mass spectrometry (MS/MS) data sets were searched against several databases: the complete set of *P. y. yoelii* full and partial proteins (7,860 total); 791,324 *P. y. yoelii* open reading frames (stop-to-stop ORFs over 15 amino acids and start-to-stop ORFs over 100 amino acids); 57,885 ORFs from NCBI's RefSeq for human, mouse and rat; 15,570 *Anopheles*, *Aedes* and *Drosophila melanogaster* proteins from GenBank; and 165 common protein contaminants (for example, trypsin, bovine serum albumin).

Gene finding and annotation

The splice site recognition module of GlimmerMExon was trained specifically for *P. yoelii* genome data, using DNA sequences extracted from a set of 1,166 donor and 1,166 acceptor sites confirmed by *P. y. yoelii* ESTs. Phat and the exon recognition module of GlimmerMExon were trained on *P. falciparum* data as described (see ref. 54). Combiner was used to generate a final ranked list of *P. y. yoelii* gene models, and TIGR's Eukaryotic Genome Control suite of programs was used for automated annotation of these (both described in ref. 54). Automated gene names were assigned to proteins by taking the 'equivalence' name of the hidden Markov model (HMM) associated with the protein where possible, or where no HMM was assigned, on the basis of the best-paired alignment. Each protein was assigned an identifier in the format 'PY000001'.

Paralogous gene families

Proteins encoded by multigene families were identified by a domain-based clustering algorithm developed at TIGR. Families were regarded as potentially *Plasmodium*- or *yoelii*-specific if they were not described by any Pfam⁵⁵ or TIGRFAM⁵⁶ domains and if the automatic annotation process had not ascribed names corresponding to widely distributed proteins. HMMs for these families were built using the HMMER package version 2.1.1 (ref. 57). Newly constructed models were then used to search the *P. yoelii*, *P. falciparum* and GenBank databases to define the scope of the families.

Telomeric/subtelomeric repeat analysis

Subtelomeric contigs were identified through alignment using MUMmer² (ref. 40) with a minimum exact match ranging from 30–40 bases. Tandem Repeat Finder⁵⁸ used the following settings: match = 2, mismatch = 7, PM (match probability) = 75, PI (indel probability) = 10, minscore = 400, max period = 700.

Comparative analyses

Gene model predictions in the syntenic region of *P. falciparum* chromosome 7 were inspected manually, and bi-directional best hits between gene models that respected conserved synteny were selected. A global alignment of the two sequences was calculated using Owen⁵⁹, and nucleotide sequences of predicted gene models were aligned using CLUSTALW⁶⁰ with default parameters, and refined manually. The number of substitutions per synonymous (d_s) and nonsynonymous (d_n) sites were estimated using the Nei and Gojobori method⁶¹. Conservation of gene order was established using Position Effect (<http://www.tigr.org/software>), where matches between *P. falciparum* and *P. y. yoelii* genes were calculated using BLASTP with a cutoff E value of 10^{-15} . The query and hit gene from each match were defined as anchor points in gene sets composed of adjacent genes. Up to ten genes upstream and downstream from each anchor gene were used in creating the gene set. An optimal alignment was calculated between the ordered gene sets using BLASTP per cent similarity scores and a linear gap penalty. Low-scoring alignments with a cumulative per cent similarity less than 100 were not used. Each optimal alignment provided a list of matching genes in conserved order between *P. falciparum* and *P. y. yoelii*.

Received 31 July; accepted 30 August 2002; doi:10.1038/nature01099.

- Carter, R. & Diggs, C. L. *Parasitic Protozoa* 359–465 (Academic, New York/San Francisco/London, 1977).
- van Dijk, M. R., Waters, A. P. & Janse, C. J. Stable transfection of malaria parasite blood stages. *Science* **268**, 1358–1362 (1995).
- Carlton, J. M. & Carucci, D. J. Rodent models of malaria in the genomics era. *Trends Parasitol.* **18**, 100–102 (2002).
- Myers, E. W. *et al.* A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
- Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
- McCutchan, T. F., Dame, J. B., Miller, L. H. & Barnwell, J. Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science* **225**, 808–811 (1984).
- Sutton, G. G., White, O., Adams, M. D. & Kervallage, A. R. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci. Technol.* **1**, 9–19 (1995).
- Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**, 1661–1671 (2002).
- Janse, C. J., Carlton, J. M.-R., Walliker, D. & Waters, A. P. Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol. Biochem. Parasitol.* **68**, 285–296 (1994).
- Daly, T. M., Long, C. A. & Bergman, L. W. Interaction between two domains of the *P. yoelii* MSP-1 protein detected using the yeast two-hybrid system. *Mol. Biochem. Parasitol.* **117**, 27–35 (2001).

- Quackenbush, J. *et al.* The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.* **29**, 159–164 (2001).
- Cawley, S. E., Wirth, A. I. & Speed, T. P. Phat—a gene finding program for *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **118**, 167–174 (2001).
- Salzberg, S. L., Pertea, M., Delcher, A. L., Gardner, M. J. & Tettelin, H. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59**, 24–31 (1999).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Thompson, J., Janse, C. J. & Waters, A. P. Comparative genomics in *Plasmodium*: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol.* **118**, 147–154 (2001).
- Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
- Suroليا, N. & Suroليا, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nature Med.* **7**, 167–173 (2001).
- Ohkanda, J. *et al.* Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. *Bioorg. Med. Chem. Lett.* **11**, 761–764 (2001).
- The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res.* **11**, 1425–1433 (2001).
- Cooke, B. M., Mohandas, N. & Coppel, R. L. The malaria-infected red blood cell: structural and functional changes. *Adv. Parasitol.* **50**, 1–86 (2001).
- del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842 (2001).
- Janssen, C. S., Barrett, M. P., Turner, C. M. & Phillips, R. S. A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proc. R. Soc. Lond. B* **269**, 431–436 (2002).
- Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
- Preiser, P. R., Jarra, W., Capiod, T. & Snounou, G. A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* **398**, 618–622 (1999).
- Galinski, M. R., Xu, M. & Barnwell, J. W. *Plasmodium vivax* reticulocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the *Plasmodium yoelii* 235 kDa rhoptry protein family. *Mol. Biochem. Parasitol.* **108**, 257–262 (2000).
- Rayner, J. C., Galinski, M. R., Ingravallo, P. & Barnwell, J. W. Two *Plasmodium falciparum* genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proc. Natl Acad. Sci. USA* **97**, 9648–9653 (2000).
- Wiser, M. F., Giraldo, L. E., Schmitt-Wrede, H. P. & Wunderlich, F. *Plasmodium chabaudi*: immunogenicity of a highly antigenic glutamate-rich protein. *Exp. Parasitol.* **85**, 43–54 (1997).
- Spielmann, T. & Beck, H. P. Analysis of stage-specific transcription in *Plasmodium falciparum* reveals a set of genes exclusively transcribed in ring stage parasites. *Mol. Biochem. Parasitol.* **111**, 453–458 (2000).
- Favaloro, J. M., Culvenor, J. G., Anders, R. F. & Kemp, D. J. A *Plasmodium chabaudi* antigen located in the parasitophorous vacuole membrane. *Mol. Biochem. Parasitol.* **62**, 263–270 (1993).
- Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
- Scherf, A., Figueiredo, L. M. & Freitas-Junior, L. H. *Plasmodium* telomeres: a pathogen's perspective. *Curr. Opin. Microbiol.* **4**, 409–414 (2001).
- Mu, J. *et al.* Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* **418**, 323–326 (2002).
- Garnham, P. C. C. *Malaria Parasites and Other Haemosporidia* (Blackwell Scientific, Oxford, 1966).
- Escalante, A. A. & Ayala, F. J. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl Acad. Sci. USA* **91**, 11373–11377 (1994).
- Waters, A. P., Higgins, D. G. & McCutchan, T. F. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc. Natl Acad. Sci. USA* **88**, 3140–3144 (1991).
- Tchavtchitch, M., Fischer, K., Huestis, R. & Saul, A. The sequence of a 200 kb portion of a *Plasmodium vivax* chromosome reveals a high degree of conservation with *Plasmodium falciparum* chromosome 3. *Mol. Biochem. Parasitol.* **118**, 211–222 (2001).
- Carlton, J. M.-R., Galinski, M. R., Barnwell, J. W. & Dame, J. B. Karyotype and synteny among the chromosomes of all four species of human malaria parasite. *Mol. Biochem. Parasitol.* **101**, 23–32 (1999).
- Janse, C. J. Chromosome size polymorphism and DNA rearrangements in *Plasmodium*. *Parasitol. Today* **9**, 19–22 (1993).
- Carlton, J. M. R., Vinkenoog, R., Waters, A. P. & Walliker, D. Gene synteny in species of *Plasmodium*. *Mol. Biochem. Parasitol.* **93**, 285–294 (1998).
- Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).
- van Lin, L. H. M., Pace, T., Janse, C. J., Scotti, R. & Ponzi, R. A long range restriction map of chromosomes 5 of *Plasmodium berghiei* demonstrates a chromosomespecific symmetrical subtelomeric organisation. *Mol. Biochem. Parasitol.* **86**, 111–115 (1997).
- Janse, C. J., Ramesar, J., van den Berg, F. M. & Mons, B. *Plasmodium berghiei*: in vivo generation and selection of karyotype mutants and non-gametocyte producer mutants. *Exp. Parasitol.* **74**, 1–10 (1992).
- Liu, S. L. & Sanderson, K. E. Rearrangements in the genome of the bacterium *Salmonella typhi*. *Proc. Natl Acad. Sci. USA* **92**, 1018–1022 (1995).
- Dame, J. B. & McCutchan, T. F. The four ribosomal DNA units of the malaria parasite *Plasmodium berghiei*. Identification, restriction map and copy number analysis. *J. Biol. Chem.* **258**, 6984–6990 (1983).
- van Lin, L. H. *et al.* Interspecies conservation of gene order and intron-exon structure in a genomic locus of high gene density and complexity in *Plasmodium*. *Nucleic Acids Res.* **29**, 2059–2068 (2001).
- Jareborg, N., Birney, E. & Durbin, R. Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res.* **9**, 815–824 (1999).
- Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A. & Kondrashov, A. S. Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* **17**, 373–376 (2001).
- Makalowski, W. & Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl Acad. Sci. USA* **95**, 9407–9412 (1998).
- Carlton, J. M., Fidock, D. A., Djimde, A., Plowe, C. V. & Welles, T. E. Conservation of a novel

- vacuolar transporter in *Plasmodium* species and its central role in chloroquine resistance of *P. falciparum*. *Curr. Opin. Microbiol.* **4**, 415–420 (2001).
50. Weinbaum, F. I., Evans, C. B. & Tigelaar, R. E. An *in vitro* assay for T cell immunity to malaria in mice. *J. Immunol.* **116**, 1280–1283 (1976).
 51. Landau, I. & Chabaud, A. G. Natural infection by 2 plasmodia of the rodent *Thamnomys rutilans* in the Central African Republic. *C.R. Acad. Sci. Hebd. Seances Acad. Sci. D* **261**, 230–232 (1965).
 52. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
 53. Beetsma, A. L., van de Wiel, T. J., Sauerwein, R. W. & Eling, W. M. *Plasmodium berghei* ANKA: purification of large numbers of infectious gametocytes. *Exp. Parasitol.* **88**, 69–72 (1998).
 54. Gardner, M. J. *et al.* Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* **419**, 531–534 (2002).
 55. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
 56. Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–43 (2001).
 57. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
 58. Benson, G. Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
 59. Ogurtsov, A. Y., Roytberg, M. A., Shabalina, S. A. & Kondrashov, A. S. OWEN: aligning long collinear regions of genomes. *Bioinformatics* (in the press).
 60. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
 61. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
 62. Black, C. G., Wang, L., Hibbs, A. R., Werner, E. & Coppel, R. L. Identification of the *Plasmodium chabaudi* homologue of merozoite surface proteins 4 and 5 of *Plasmodium falciparum*. *Infect. Immun.* **67**, 2075–2081 (1999).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank S. Cawley and T. Pace for collaborative work; J. Mendoza and J. Ramesar for technical support; C. Long for the gift of a *P. y. yoelii* cDNA library; R. Arcilla and W. Weiss for parasite material; and J. Eisen and S. Sullivan for critical reading of the manuscript. L.H.v.L. was supported by an INCO-DEV programme grant from the European Community; T.W.K. was supported by a Rijks Universiteit te Leiden studentship; J.D.R. was supported with funds from the Wellcome Trust. This project was funded by the US Department of Defense through cooperative agreement with the US Army Medical Research and Material Command and by the Naval Medical Research Center. The opinions expressed are those of the authors and do not reflect the official policy of the Department of the Navy, Department of Defense, or the US government.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.C. (e-mail: carlton@tigr.org). Access to genome annotation data is available through the TIGR Eukaryotic Projects website (<http://www.tigr.org>) and PlasmoDB (<http://www.plasmodb.org>). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the project accession number AABL00000000. The version described in this paper is the first version, AABL01000000.

A proteomic view of the *Plasmodium falciparum* life cycle

Laurence Florens*, Michael P. Washburn†, J. Dale Raine‡, Robert M. Anthony§, Munira Grainger||, J. David Haynes¶, J. Kathleen Moch§, Nemone Muster*, John B. Sacchi§#, David L. Tabb*☆, Adam A. Witney§#, Dirk Wolters†#, Yimin Wu**, Malcolm J. Gardner††, Anthony A. Holder||, Robert E. Sinden‡, John R. Yates*† & Daniel J. Carucci§

* Department of Cell Biology, The Scripps Research Institute, SR-11, 10550 North Torrey Pines Road, La Jolla, California 92037, USA
† Department of Proteomics and Metabolomics, Torrey Mesa Research Institute, Syngenta Research & Technology, 3115 Merryfield Row, San Diego, California 92121-1125, USA
‡ Infection and Immunity Section, Department of Biological Sciences, Imperial College of Science, Technology & Medicine, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK
§ Naval Medical Research Center, Malaria Program (IDD), 503 Robert Grant Avenue, Room 3A40; and ¶ Department of Immunology, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910-7500, USA
|| The Division of Parasitology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
☆ Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA
** Malaria Research and Reference Reagent Resource Center, American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, USA
†† The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

The completion of the *Plasmodium falciparum* clone 3D7 genome provides a basis on which to conduct comparative proteomics studies of this human pathogen. Here, we applied a high-throughput proteomics approach to identify new potential drug and vaccine targets and to better understand the biology of this complex protozoan parasite. We characterized four stages of the parasite life cycle (sporozoites, merozoites, trophozoites and gametocytes) by multidimensional protein identification technology. Functional profiling of over 2,400 proteins agreed with the physiology of each stage. Unexpectedly, the antigenically variant proteins of *var* and *rif* genes, defined as molecules on the surface of infected erythrocytes, were also largely expressed in sporozoites. The detection of chromosomal clusters encoding co-expressed proteins suggested a potential mechanism for controlling gene expression.

The life cycle of *Plasmodium* is extraordinarily complex, requiring specialized protein expression for life in both invertebrate and vertebrate host environments, for intracellular and extracellular survival, for invasion of multiple cell types, and for evasion of host immune responses. Interventional strategies including anti-malarial vaccines and drugs will be most effective if targeted at specific parasite life stages and/or specific proteins expressed at these stages. The genomes of *P. falciparum*¹ and *P. yoelii yoelii*² are now completed and offer the promise of identifying new and effective drug and vaccine targets.

Functional genomics has fundamentally changed the traditional gene-by-gene approach of the pre-genomic era by capitalizing on the success of genome sequencing efforts. DNA microarrays have been successfully used to study differential gene expression in the abundant blood stages of the *Plasmodium* parasite^{3,4}. However, transcriptional analysis by DNA microarrays generally requires microgram quantities of RNA and has been restricted to stages that can be cultivated *in vitro*, limiting current large-scale gene expression analyses to the blood stages of *P. falciparum*. As several key stages of the parasite life cycle, in particular the pre-erythrocytic stages, are not readily accessible to study, and as differential gene expression is in fact a surrogate for protein expression, global proteomic analyses offer a unique means of determining not only protein expression, but also subcellular localization and post-translational modifications.

We report here a comprehensive view of the protein complements isolated from sporozoites (the infectious form injected by the mosquito), merozoites (the invasive stage of the erythrocytes),

trophozoites (the form multiplying in erythrocytes), and gametocytes (sexual stages) of the human malaria parasite *P. falciparum*. These proteomes were analysed by multidimensional protein identification technology (MudPIT), which combines in-line, high-resolution liquid chromatography and tandem mass spectrometry⁵. Two levels of control were implemented to differentiate parasite from host proteins. By using combined host–parasite sequence databases and noninfected controls, 2,415 parasite proteins were confidently identified out of thousands of host proteins; that is, 46% of all gene products were detected in four stages of the *Plasmodium* life cycle (Supplementary Table 1).

Comparative proteomics throughout the life cycle

The sporozoite proteome appeared markedly different from the other stages (Table 1). Almost half (49%) of the sporozoite proteins

Table 1 Comparative summary of the protein lists for each stage

Protein count	Sporozoites	Merozoites	Trophozoites	Gametocytes
152	X	X	X	X
197	–	X	X	X
53	X	–	X	X
28	X	X	–	X
36	X	X	X	–
148	–	–	X	X
73	–	X	–	X
120	X	–	–	X
84	–	X	X	–
80	X	–	X	–
65	X	X	–	–
376	–	–	–	X
286	–	–	X	–
204	–	X	–	–
513	X	–	–	–
2,415	1,049	839	1,036	1,147

Whole-cell protein lysates were obtained from, on average, 17 × 10⁶ sporozoites, 4.5 × 10⁹ trophozoites, 2.75 × 10⁹ merozoites, and 6.5 × 10⁹ gametocytes.

Present addresses: BRB 13-009, Department of Microbiology and Immunology, University of Maryland School of Medicine, 655 W. Baltimore St., Baltimore, Maryland 21201, USA (J.B.S.); Department of Medical Microbiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK (A.A.W.); and Ruhr-University Bochum, Institute of Analytical Chemistry, 44780 Bochum, Germany (D.W.).

were unique to this stage, which shared an average of 25% of its proteins with any other stage. On the other hand, trophozoites, merozoites and gametocytes had between 20% and 33% unique proteins, and they shared between 39% and 56% of their proteins. Consequently, only 152 proteins (6%) were common to all four stages. Those common proteins were mostly housekeeping proteins such as ribosomal proteins, transcription factors, histones and cytoskeletal proteins (Supplementary Table 1). Proteins were sorted into main functional classes based on the Munich Information Centre for Protein Sequences (MIPS) catalogue⁶, with some adaptations for classes specific to the parasite, such as cell surface and apical organelle proteins (Fig. 1). When considering the annotated proteins in the database, some marked differences appeared between sporozoites and blood stages (Fig. 1). Although great care was taken to ensure that the results reflect the state of the parasite in the host, a portion of the data set may reflect the parasite's response to different purification treatments. However, the stage-specific detection of known protein markers at each stage established the relevance of our data set.

The merozoite proteome

Merozoites are released from an infected erythrocyte, and after a short period in the plasma, bind to and invade new erythrocytes. Proteins on the surface and in the apical organelles of the merozoite mediate cell recognition and invasion in an active process involving an actin-myosin motor. Four putative components of the invasion motor⁷, merozoite cap protein-1 (MCP1), actin, myosin A, and myosin A tail domain interacting protein (MTIP), were abundant merozoite proteins (Supplementary Table 2). Abundant merozoite surface proteins (MSPs) such as MSP1 and MSP2 are linked by a glycosylphosphatidyl (GPI) anchor to the membrane, and both have been implicated in immune evasion (reviewed in ref. 8). A second family of peripheral membrane proteins, represented by MSP3 and MSP6, was also detected (Fig. 2a), although these proteins are largely soluble proteins of the parasitophorous vacuole, which are released on schizont rupture. Other vacuolar proteins, such as the acidic basic repeat antigen (ABRA) and serine repeat antigen (SERA), were detected in the merozoite fraction, but some such as S-antigen⁹ were not (Supplementary Table 2). Notably, MSP8 and a related MSP8-like protein were only identified in sporozoites (Fig. 2a). Some MSPs are diverse in sequence and may be extensively modified by proteolysis; these features, together with the association of a variety of peripheral and soluble proteins, provide for a complex surface architecture.

Many apical organellar proteins, in the micronemes and rhoptries, have a single transmembrane domain. Among these proteins, apical membrane antigen 1 (AMA1) and MAEBL were found in

both sporozoite and merozoite preparations (Fig. 2a). Erythrocyte-binding antigens (EBA), such as EBA 175 and EBA 140/BAEBL, were found only in the merozoite and trophozoite fractions. Of note, the reticulocyte-binding protein (PFRH) family (PFD0110w, MAL13P1.176, PF13_01998, PFL2520w and PFD1150c), which has similarity with the Py235 family of *P. y. yoelii* rhoptry proteins and the *Plasmodium vivax* reticulocyte-binding proteins, was not detected in the merozoite fraction. Some PFRH proteins were, however, detected in sporozoites (Fig. 2a), including RH3, which is a transcribed pseudogene in blood stages¹⁰. Components of the low molecular mass rhoptry complex, the rhoptry-associated proteins (RAP) 1, 2 and 3, were all found in merozoites. RAP1 was also detected in sporozoites. The high molecular mass rhoptry protein complex (RhopH), together with ring-infected erythrocyte surface antigen (RESA), which is a component of dense granules, is transferred intact to new erythrocytes at or after invasion and may contribute to the host cell remodelling process. RhopH1, RhopH2 (PF11445w; Ling, I. T., *et al.*, unpublished data) and RhopH3 were found in the merozoite proteome. RhopH1 (PFC0120w/PFC0110w) has been shown to be a member of the cyto-adherence linked asexual gene family (CLAG)¹¹; however, the presence of CLAG9 in the merozoite fraction (Fig. 2a) suggests that CLAG9 may also be a RhopH protein, casting some doubt on the proposed role for this protein in cyto-adherence¹².

The trophozoite proteome

After erythrocyte invasion the parasite modifies the host cell. The principal modifications during the initial trophozoite phase (lasting about 30 h) allow the parasite to transport molecules in and out of the cell, to prepare the surface of the red blood cell to mediate cyto-adherence, and to digest the cytoplasmic contents, particularly haemoglobin, in its food vacuole. In the next phase of schizogony (the final ~18 h of the asexual development in the blood cell), nuclear division is followed by merozoite formation and release.

Knob-associated histidine-rich protein (KAHRP) and erythrocyte membrane proteins 2 and 3 (EMP2 and -3) bind to the erythrocyte cytoskeleton (Fig. 2a). Of the proteins of the parasitophorous vacuole and the tubovesicular membrane structure extending into the cytoplasm of the red blood cell, three (the skeleton-binding protein 1, and exported proteins EXP1 and EXP2) were represented by peptides (Fig. 2a); although a fourth (Sar1 homologue, small GTP-binding protein; PFD0810w) was not. It is likely that one or more of the hypothetical proteins detected only in the trophozoite sample are involved in these unusual structures.

Digestion of haemoglobin is a major parasite catabolic process¹³. Members of the plasmepsin family (aspartic proteinases; PF14_0075 to PF14_0078)¹⁴, falcipain family (cysteine proteinases; PF11_0161,

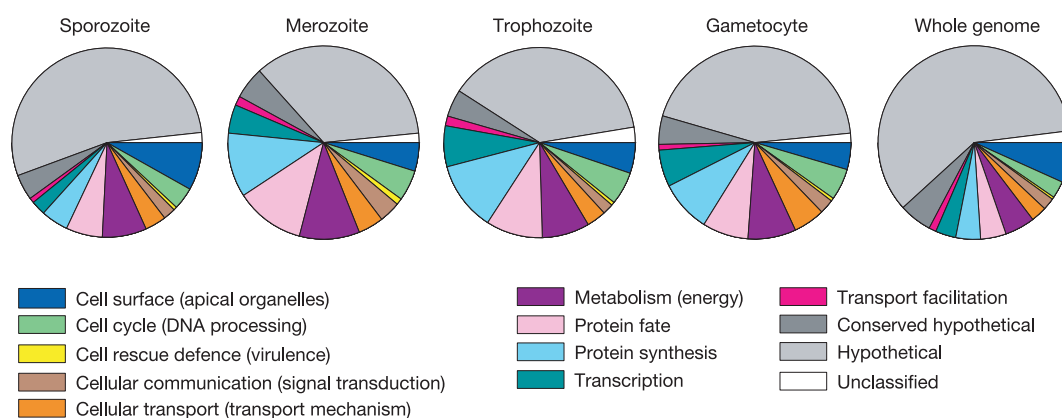


Figure 1 Functional profiles of expressed proteins. Proteins identified in each stage are plotted as a function of their broad functional classification as defined by the MIPS

catalogue⁶. To avoid redundancy, only one class was assigned per protein. The complete protein list is given in Supplementary Table 1.

PF11_0162 and PF11_0165)¹⁵, and falcilysin (a metallopeptidase; PF13_0322)¹⁶ implicated in this process were all clearly identified (Supplementary Table 1). Several proteases expressed in the merozoite and trophozoite fractions, and not involved in haemoglobin digestion, may be important in parasite release at the end of schizogony, invasion of the new cell, or merozoite protein processing. Possible candidates for this mechanism include cysteine proteinases of the falcipain and SERA families, or subtilisins such as SUB1 and SUB2, both located in apical organelles (Fig. 2a).

The gametocyte proteome

Stage V gametocytes are dimorphic, with a male:female ratio of 1:4. They are arrested in the cell cycle until they enter the mosquito where development is induced within minutes to form the male and

female gametes. Gametocyte structure reflects these ensuing fates; that is, the female has abundant ribosomes and endoplasmic reticulum/vesicular network to re-initiate translation, whereas the male is largely devoid of ribosomes and is terminally differentiated¹⁷.

Gametocyte-specific transcription factors, RNA-binding proteins, and gametocyte-specific proteins involved in the regulation of messenger RNA processing (particularly splicing factors, RNA helicases, RNA-binding proteins, ribonucleoproteins (RNPs) and small nuclear ribonucleoprotein particles (snRNPS)) were highly represented in the gametocyte proteome (Supplementary Table 1). Transcription in the terminally differentiated gametocytes is ‘suppressed’, but the female gametocytes contain mRNAs encoding gamete/zygote/ookinete surface antigens (for example, P25/28)

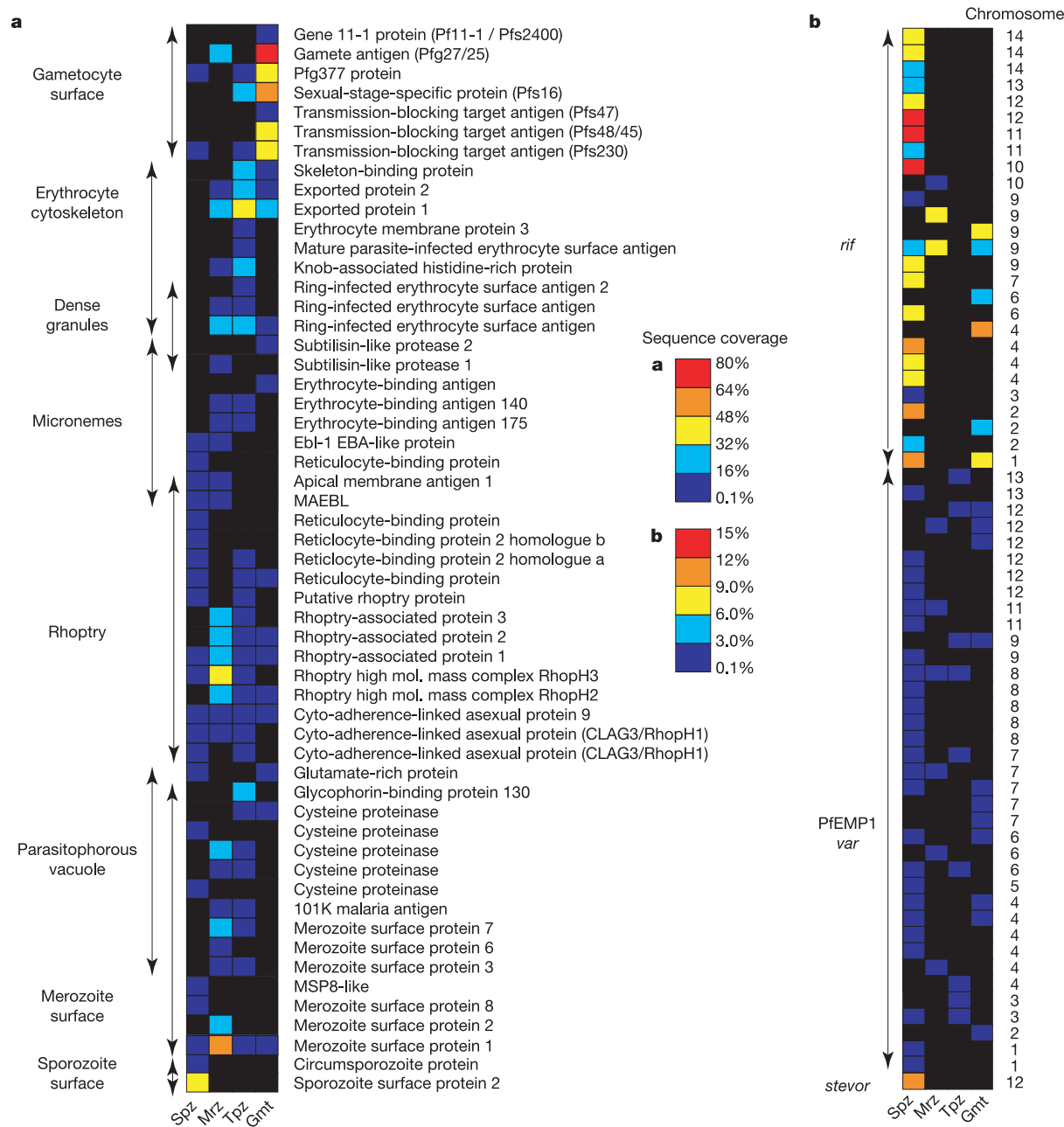


Figure 2 Expression patterns of known stage-specific proteins. **a**, Cell surface, organelle, and secreted proteins are plotted as a function of their known subcellular localization. **b**, *stevor*, *var* and *rif* polymorphic surface variants are plotted as a function of the chromosome encoding their genes. The matrices are colour-coded by sequence coverage

measured in each stage (proteins not detected in a stage are represented by black squares). Locus names associated with these proteins are listed in Supplementary Table 2. Spz, sporozoite; mrz, merozoite; tpz, trophozoite; gmt, gametocyte.

that are subject to post-transcriptional control; this control is released rapidly during gamete development¹⁷. Ribosomal proteins were largely represented: 82% of known small subunit (SSU) proteins and 69% of known large subunit (LSU) proteins were detected in gametocytes compared to 94% and 82%, respectively, from all stages examined (Supplementary Table 1). We suggest that this reflects the accumulation of ribosomes in the female gametocyte to accommodate for the sudden increase in protein synthesis required during gametogenesis and early zygote development.

Other protein groupings highly represented in the gametocyte were in the cell cycle/DNA processing and energy classes (Fig. 1). The former is consistent with the biological observation that the mature gametocyte is arrested in G0 of the cell cycle and will require a full complement of pre-existing cell cycle regulatory cascades to respond, within seconds, to the gametogenesis stimuli (that is, xanthurenic acid and a drop in temperature)¹⁸. Metabolic pathways of the malaria parasite may be stage-specific, with asexual blood stage parasites dependent on glycolysis and conversion of pyruvate to lactate (L-lactate dehydrogenase) for energy. In the gametocyte and sporozoite preparations, peptides from enzymes involved in the mitochondrial tricarboxylic acid (TCA) cycle and oxidative phosphorylation were identified (Table 2). This observation suggests that gametocytes have fully functional mitochondria as a pre-adaptation to life in the mosquito, as suggested by morphological and biochemical studies¹⁹ and their sensitivity to anti-malarials attacking respiration (primaquine and artesinin-based products)¹⁷. It will be interesting to observe whether other mosquito and liver stages, which show similar drug sensitivities, express the same metabolic proteome.

Cell surface proteins (Fig. 1) included most of the known surface antigens (Fig. 2a and Supplementary Table 2). However, Pfs35 and a sexual stage-specific kinase (PF13_0258) were not detected. Nevertheless the cultured gametocytes analysed in this study expressed a specific repertoire of rifin and PfEMP1 proteins (Fig. 2b and Supplementary Table 2). Together these observations suggest that the gametocyte, which is very long-lived in the red blood cell (that is, 9–12 days compared with 2 days for the pathogenic asexual parasites), expresses a limited repertoire of the highly polymorphic families of surface antigens so widely represented in the asexual parasites.

The sporozoite proteome

Sporozoites are injected by the mosquito during ingestion of a blood meal. Although, they are in the blood stream for only minutes, sporozoites probably require mechanisms to evade the host humoral immune system in order for at least a fraction of the thousands of sporozoites injected by the mosquito to survive the

hostile environment in the blood and successfully invade hepatocytes.

The main class of annotated sporozoite proteins identified was cell surface and organelle proteins (Fig. 1). Sporozoites are an invasive stage and possess the apical complex machinery involved in host cell invasion. As observed in the analysis of the *P. y. yoelii* sporozoite transcriptome²⁰, actin and myosin were found in the motile sporozoites (Supplementary Table 2). Many proteins associated with rhoptry, micronemes and dense granules were detected (Fig. 2a). Among the proteins found were known markers of the sporozoite stage, such as the circumsporozoite protein (CSP) and sporozoite surface protein 2 (SSP2; also known as TRAP), both present in large quantities at the sporozoite surface (Fig. 2a). Peptides derived from CTRP (circumsporozoite protein and thrombospondin-related adhesive protein (TRAP)-related protein), an ookinete cell surface protein involved in recognition and/or motility²¹, were detected in the sporozoite fractions (Supplementary Table 1).

Most surprisingly, peptides derived from multiple *var* (coding for PfEMP1) and *rif* genes were identified in the sporozoite samples. PfEMP1 and rifins are coded for by large multigene families (*var* and *rif*)^{22,23} and are present on the surface of the infected red blood cell. No peptides derived from *rif* genes were identified in the trophozoite sample, whereas sporozoites expressed 21 different rifins and 25 PfEMP1 isoforms (Fig. 2b); that is, a total of 14% of the *rif* genes and 33% of the *var* genes encoded by the genome. Furthermore, very little overlap was observed between stages: only ten PfEMP1 and two rifin isoforms expressed in sporozoites were found in other stages. Whereas in the blood stream the asexual stage parasites undergo asexual multiplication and therefore have an opportunity to undergo antigenic 'switching' of the variant antigen genes, the non-replicative sporozoites may not have this opportunity. Expressing such a polymorphic array of *var* (PfEMP1) and *rif* genes could be part of a sporozoite survival mechanism.

Chromosomal clusters encoding co-expressed proteins

The distinct proteomes of each stage of the *Plasmodium* life cycle suggested that there is a highly coordinated expression of *Plasmodium* genes involved in common processes. Co-expression groups are a widespread phenomenon in eukaryotes, where mRNA array analyses have been used to establish gene expression profiles. Analysis of co-regulated gene groups facilitates both searching for regulatory motifs common to co-regulated genes, and predicting protein function on the basis of the 'guilt by association' model. Furthermore, mRNA analyses in *Saccharomyces cerevisiae*²⁴ and *Homo sapiens*^{25,26} have demonstrated that co-regulated genes do not map to random locations in the genome but are in fact

Table 2 Examples on enzymes in stage-specific metabolic pathways

Locus	Stage				Enzyme	EC number†	Reaction catalysed
	Spz*	Mrz*	Tpz*	Gmt*			
End of glycolysis							
PF10_0363	1.2	–	2.4	–	Pyruvate kinase	2.7.1.40	P-enolpyruvate to pyruvate
MAL6P1.160	8.6	66.9	18.8	14.7	Pyruvate kinase		
PF13_0141	46.2	83.9	70.9	78.8	L-lactate dehydrogenase	1.1.1.27	Pyruvate to lactate
TCA cycle and oxidative phosphorylation							
PF10_0218	12.3	–	–	–	Citrate synthase	4.1.3.7	Acetyl CoA + oxaloacetate to citrate
PF13_0242	3.2	–	16.9	8.8	Isocitrate dehydrogenase (NADP)	1.1.1.41	Isocitrate to 2-oxoglutarate + CO ₂
PF08_0045	2.9	–	2.2	23.1	2-Oxoglutarate dehydrogenase e1 component	1.2.4.2	2-Oxoglutarate to succinyl CoA
PF10_0334	–	–	3.5	27.7	Flavoprotein subunit of succinate dehydrogenase	1.3.5.1	Succinate to fumarate
PFL0630w	3.7	–	–	12.1	Iron-sulphur subunit of succinate dehydrogenase		
PF14_0373	–	–	–	12.7	Ubiquinol cytochrome oxidoreductase	1.10.2.2	Ubiquinol to cytochrome c reductase in electron transport
PFB0795w	–	–	–	14.2	ATP synthase F1, α-subunit		
PFI1365w	–	–	–	8.8	Cytochrome c oxidase subunit	1.9.3.1	
PFI1340w	–	–	–	8.8	Fumarate hydratase	4.2.1.2	Fumarate to malate
MAL6P1.242	30.4	–	–	40.9	Malate dehydrogenase	1.1.1.37	Malate to oxaloacetate

Plasmodium metabolic pathways can be found at <http://www.sites.huji.ac.il/malaria/>. Spz, sporozoite; mrz, merozoite; tpz, trophozoite; gmt, gametocyte.

* The sequence coverage (that is, the percentage of the protein sequence covered by identified peptides) measured in each stage is reported.

† Enzyme Commission (EC) numbers are reported for each protein.

frequently organized into gene clusters on a chromosome. Gene clustering in *Plasmodium* species has been demonstrated. Ordered arrays of genes involved in virulence and antigenic variation (for example, *var*, *vir* and *rif* genes) are located in the subtelomeric regions of the chromosomes^{27,28}.

To determine whether gene clustering exists along the entire *P. falciparum* genome, genes whose protein products were detected in our analysis were mapped onto all 14 chromosomes in a stage-dependent manner (Fig. 3a). The 2,415 proteins identified represented an average of 45% of the open reading frames (ORFs) predicted per chromosome. The number of protein hits by chromosome was similar for all stages: sporozoite, merozoite, trophozoite and gametocyte protein lists constituting 19.7%, 15.8%, 19.5% and 21.6% of the predicted ORFs per chromosomes, respectively. Groups of three or more consecutive loci whose protein products were detected in a particular stage were defined as chromosomal clusters encoding co-expressed proteins (Fig. 3b). On the basis of this definition a total of 98 clusters containing 3 loci, 32 clusters containing 4 loci, 5 clusters containing 5 loci, and 3 clusters containing 6 loci were identified (Supplementary Table 3). For each chromosome, the frequency of finding clusters encoding co-expressed proteins containing 3–6 adjacent loci markedly exceeded

the probability of finding such clusters by chance (see the footnote of Supplementary Table 3 for details on the probability calculation). Therefore, chromosomal clusters encoding co-expressed proteins were prevalent in the *P. falciparum* genome.

Functionally related genes have been shown to cluster in the *S. cerevisiae*²⁴ and human genomes²⁶. This phenomenon also occurs in *P. falciparum*. A total of 138 clusters encoding co-expressed proteins were identified and 67 of them (49%) contained at least two loci that have been functionally annotated. Of these 67 clusters, 30 contained at least two loci whose annotation clearly indicates that the proteins are functionally related. For example, clusters on chromosomes 3, 5 and 10 contained ribosomal proteins, proteins involved in protein modification, and proteins involved in nucleotide metabolism, respectively (Table 3). Chromosome 14 contained a cluster of four aspartic proteases co-expressed in all of the blood stages (Table 3). This cluster was not detected in sporozoites, where no haemoglobin degradation is expected to occur. Interestingly, whereas the falcipain gene cluster on chromosome 11 appeared in our analysis as a cluster of co-expressed proteins (Supplementary Table 3), the SERA gene cluster on chromosome 2, coding for proteins that share a papain-like sequence motif²⁹, did not. Of the ten sporozoite-specific clusters, five involved *var* and *rif* genes, such as the *rif* cluster located in the subtelomeric domain of chromosome 14 (Table 3). On the basis of their presence in clusters encoding co-expressed proteins, we were able to suggest functional roles for 24 proteins annotated as hypothetical in the *P. falciparum* genome (Supplementary Table 3). For example, a gametocyte-specific cluster on chromosome 13 encoded two transmission-blocking antigens (Pfs48/45 and Pfs47) and a hypothetical protein, PF13_0246, which might be a gametocyte surface protein. Two clusters on chromosomes 2 and 11 were highly specific to the trophozoite stage (Table 3). Each of these clusters contained well-known secreted and surface proteins, namely KAHRP, PfEMP3, antigen 332, and RESA, all of which have been implicated in knob formation. The highly coordinated expression of these genes makes the three hypothetical proteins listed in these trophozoite-specific gene clusters possible candidates for involvement in cyto-adherence.

Discussion

Although sample handling is a principal consideration when studying pathogens, the expression of large numbers of previously identified proteins was consistent with their published expression profiles, validating our data set as a meaningful sampling of each stage's proteome. This is a particularly important aspect of our analysis as 65% of the 5,276 genes encoded by the *P. falciparum* genome are annotated as hypothetical¹, and of the 2,415 expressed proteins we identified, 51% are hypothetical proteins (Supplementary Table 1). Our results confirmed that these hypothetical ORFs predicted by gene modelling algorithms were indeed coding regions. Furthermore, from all four stages analysed, we identified 439 proteins predicted to have at least one transmembrane segment or a GPI addition signal (18% of the data set) and 304 soluble proteins with a signal sequence; that is, potentially secreted or located to organelles. Well over half of the secreted proteins and integral membrane proteins detected were annotated as hypothetical (Supplementary Table 4). The obvious interest in this class of proteins is that, with no homology to known proteins, they represent potential *Plasmodium*-specific proteins and may provide targets for new drug and vaccine development.

Our comprehensive large-scale analysis of protein expression showed that most surface proteins are more widely expressed than initially thought. In particular, the *var* and *rif* genes, which were thought to be involved in immune evasion only in the blood stage, have now been shown to be expressed in apparently large and varied numbers at the sporozoite stage. These surface proteins might be involved in general interaction processes with host cells and/or immune evasion. An alternative hypothesis is that stage-specific

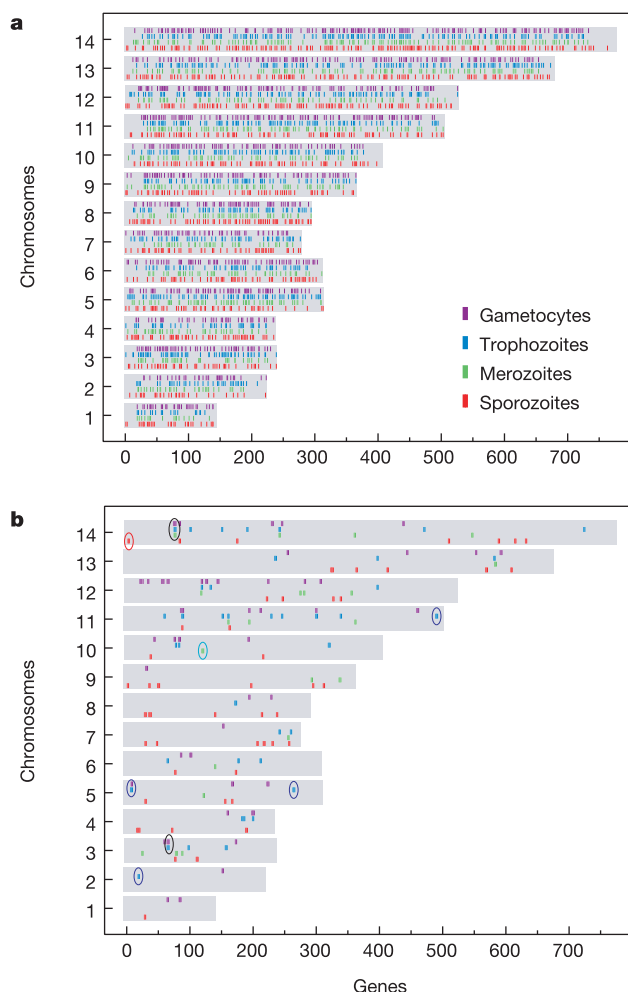


Figure 3 Distribution of expressed proteins by chromosome. **a**, For each stage, genes whose products were detected (coloured vertical bars) are plotted in the order they appear on their chromosome (grey boxes). **b**, Groups of at least three consecutive expressed genes are defined as chromosomal clusters of co-expressed proteins. Examples of such clusters, circled in **b**, are specified in Table 3 and the complete description of the 138 clusters can be found in Supplementary Table 3.

Table 3 Examples of chromosomal gene clusters encoding co-expressed proteins

Chromosome	ID	Locus	Stage				Description	Class	SP	TM
			Spz	Mrz	Tpz	Gmt				
3	64	PFC0285c	2.1	12.7	33.2	18.7	T-complex protein β -subunit	Protein fate	0	0
3	65	PFC0290w	8.3	–	33.8	18.6	40S ribosomal protein S23	Protein synthesis	0	0
3	66	PFC0295c	–	14.9	52.5	21.3	40S ribosomal protein S12	Protein synthesis	0	0
3	67	PFC0300c	–	12.1	30.4	17.9	60S ribosomal protein L7	Protein synthesis	0	0
5	263	PFE1345c	–	–	1.9	1.6	Minichromosome maintenance protein 3	Cell transport	0	0
5	264	PFE1350c	–	–	22.4	–	Ubiquitin-conjugating enzyme	Protein fate	0	0
5	265	PFE1355	–	4.8	2.6	2.6	Ubiquitin carboxy-terminal hydrolase	Protein fate	0	0
5	266	PFE1360c	–	–	7.7	–	Methionine aminopeptidase	Protein fate	0	0
10	119	PF10_0121	10.8	74.5	29	–	Hypoxanthine phosphoribosyltransferase	Metabolism	0	0
10	120	PF10_0122	5.4	6.1	–	6.1	Phosphoglucosyltransferase	Metabolism	0	0
10	121	PF10_0123	–	11.7	–	–	GMP synthetase	Metabolism	0	0
10	122	PF10_0124	0.9	1.8	–	–	Hypothetical protein	–	0	0
14	74	PF14_0074	26.6	–	–	4.9	Hypothetical protein	–	0	0
14	75	PF14_0075	–	26.5	43.2	47.4	Plasmepsin	Protein fate	1	0
14	76	PF14_0076	–	6.6	35.2	10	Plasmepsin 1	Protein fate	1	0
14	77	PF14_0077	–	21.2	43	11.5	Plasmepsin 2	Protein fate	1	0
14	78	PF14_0078	–	14.2	52.8	29.9	HAP protein	Protein fate	1	0
14	2	PF14_0002	3.5	–	–	–	Rifin	Surface or organelles	0	1
14	3	PF14_0003	7.9	–	–	–	Rifin	Surface or organelles	1	2
14	4	PF14_0004	6.5	–	–	–	Rifin	Surface or organelles	1	2
2	18	PFB0090c	–	–	3	–	Hypothetical protein, conserved	–	0	0
2	19	PFB0095c	–	–	3.4	–	Erythrocyte membrane protein 3	Surface or organelles	1	0
2	20	PFB0100c	–	1.5	24.8	–	Knob-associated histidine-rich protein	Surface or organelles	1	0
11	489	PF11_0506	–	–	6.3	4.4	Hypothetical protein	–	0	1
11	490	PF11_0507	–	–	0.8	–	Antigen 332	Surface or organelles	0	0
11	491	PF11_0508	–	–	3.3	–	Hypothetical protein	–	0	0
11	492	PF11_0509	–	6.4	3	–	RESA	Surface or organelles	0	0
13	443	PF13_0246	4.5	–	–	8.6	Hypothetical protein	–	0	0
13	444	PF13_0247	–	–	–	32.4	Transmission-blocking target antigen precursor (Pfs48/45)	Surface or organelles	1	1
13	445	PF13_0248	–	–	–	7.1	Transmission-blocking target antigen precursor (Pfs47)	Surface or organelles	1	1

Clusters of at least three consecutive genes encoding co-expressed proteins are reported with their position (ID) on the chromosome, the sequence coverage measured for these proteins in each stage (%), their current annotation and functional class, and the predicted presence of signal peptide (SP) or transmembrane domains (TM) (based on the TMHMM⁴³, a transmembrane (TM) helices prediction method based on a hidden Markov model (HMM), big-PI Predictor⁴⁴ and SignalP⁴⁵ algorithms).

regulation is not as exact as previously thought.

One mechanism of protein expression control that contributes to stage specificity in *P. falciparum* arises from the chromosomal clustering of genes encoding co-expressed proteins. The clusters described in this study demonstrate a widespread high order of chromosomal organization in *P. falciparum* and probably correspond to regions of open chromatin allowing for co-regulated gene expression. The high (A + T) content of the *P. falciparum* genome makes the identification of regulatory sequences such as promoters and enhancers challenging^{31,32}. Focusing analyses on stage-specific and multi-stage clusters will facilitate finding stage-specific and general *cis*-acting sequences in the *Plasmodium* genome and will help decipher gene expression regulation during the parasite life cycle.

The malaria parasite is a complex multi-stage organism, which has co-evolved in mosquitoes and vertebrates for millions of years. Designing drugs or vaccines that substantially and persistently interrupt the life cycle of this complex parasite will require a comprehensive understanding of its biology. The *P. falciparum* genome sequence and comparative proteomics approaches may initiate new strategies for controlling the devastating disease caused by this parasite. □

Methods

Parasite material

Plasmodium falciparum clone 3D7 (Oxford) was used throughout. Sporozoites were initially isolated from the salivary glands of *Anopheles stephensi* mosquitoes, 14 days after infection, by centrifugation in a Renograffin 60 gradient, as described³³. Four sporozoite samples were used as is. A fifth sample underwent an additional purification step on Dynabeads M-450 Epoxy coupled to NFS1 (an anti-*P. falciparum* CS protein monoclonal antibody)³⁴ according to the manufacturer's instructions (Dyna). Trophozoite-infected erythrocytes from synchronized cultures were purified on 70% Percoll-alanine³⁰, and the trophozoites released from the erythrocytes³⁵. Of the 260 parasitized erythrocytes counted by Giemsa-stained thin-blood film, 100% were identified as trophozoites. Merozoites were prepared essentially as described in ref. 36, using highly synchronized

schizonts and purifying the merozoites by passage through membrane filters. Starting with synchronized asexual parasites grown in suspension culture as described^{37,38}, gametocytes were prepared by daily media changes of static cultures at 37 °C. When there were very few mature asexual stages present, gametocyte-infected erythrocytes were collected from the 52.5%/45% and 45%/30% interfaces of a Percoll gradient³⁹. The gametocytes consisted mostly of stage IV and V parasites with minor contamination (<3%) from mixed asexual stage parasites. Finally, cellular debris from the upper bodies of parasite-free *A. stephensi* and non-infected human erythrocytes were used as controls for sporozoites and blood-stage parasites, respectively. Every effort was made to minimize enzymatic activity and protein degradation during sampling, and the subsequent isolation of the parasites; however, we cannot exclude that some of the differences in protein profiles that we observe between the different life-cycle stages may be a consequence of the sample-handling procedures.

Cell lysis

Five sporozoite, four merozoite, four trophozoite and three gametocyte preparations were lysed, digested and analysed independently. Cell pellets were first diluted ten times in 100 mM Tris-HCl pH 8.5, and incubated in ice for 1 h. After centrifugation at 18,000 g for 30 min, supernatants were set aside and microsomal membrane pellets were washed in 0.1 M sodium carbonate, pH 11.6. Soluble and insoluble protein fractions were separated by centrifugation at 18,000 g for 30 min. Supernatants obtained from both centrifugation steps were either combined (sporozoites, trophozoites and merozoites) or digested and analysed independently (gametocytes).

Peptide generation and analysis

The method follows that of Washburn *et al.*⁵, with the exception that Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl; Pierce) was used to reduce urea-denatured proteins. Peptide mixtures were analysed through MudPIT as described⁵.

Protein sequence databases

The *P. falciparum* database contained 5,283 protein sequences. Spectra resulting from contaminant mosquito and erythrocyte peptides had to be taken into account in the sporozoite and blood-stage samples, respectively. Tandem mass spectrometry (MS/MS) data sets from blood stages were therefore searched against a database containing both *P. falciparum* protein sequences and 24,006 ORFs from the human, mouse and rat RefSeq NCBI databases. At the date of the searches, the *Anopheles gambiae* genome was not available. The NCBI database contained 922 *Anopheles* and 313 *Aedes* proteins, which were combined to the 14,335 ORFs of the NCBI *Drosophila melanogaster*⁴⁰ database to create a control diptera database. Finally, these databases were complemented with a set of 172 known protein contaminants, such as proteases, bovine serum albumin and human keratins.

MS/MS data set analysis

The SEQUEST algorithm was used to match MS/MS spectra to peptides in the sequence databases⁴¹. To account for carboxyamidomethylation, MS/MS data sets were searched with a relative molecular mass of 57,000 (M_r , 57K) added to the average molecular mass of cysteines. Peptide hits were filtered and sorted with DTASelect⁴². Spectra/peptide matches were only retained if they were at least half-tryptic (Lys or Arg at either end of the identified peptide) and with minimum cross-correlation scores (XCorr) of 1.8 for +1, 2.5 for +2, and 3.5 for +3 spectra and DeltaCn (top match's XCorr minus the second-best match's XCorr divided by the top match's XCorr) of 0.08. Peptide hits were deemed unambiguous only if they were not found in non-infected controls and were uniquely assigned to parasite proteins by searching against combined parasite–host databases. Finally, for low coverage loci, peptide/spectrum matches were visually assessed on two main criteria: any given MS/MS spectrum had to be clearly above the baseline noise, and both b and y ion series had to show continuity. The Contrast tool⁴² was used to compare and merge protein lists from replicate sample runs and to compare the proteomes established for the four stages.

Received 31 July; accepted 9 September 2002; doi:10.1038/nature01107.

- Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
- Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii*. *Nature* **419**, 512–519 (2002).
- Ben Mamoun, C. *et al.* Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Mol. Microbiol.* **39**, 26–36 (2001).
- Hayward, R. E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35**, 6–14 (2000).
- Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
- Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30**, 31–34 (2002).
- Pinder, J. C. *et al.* Actomyosin motor in the merozoite of the malaria parasite, *Plasmodium falciparum*: implications for red cell invasion. *J. Cell Sci.* **111**, 1831–1839 (1998).
- Holder, A. A. *Malaria Vaccine Development: a Multi-immune Response and Multi-stage Perspective* (ed. Hoffman, S. L.) 77–104 (ASM Press, Washington, 1996).
- Coppel, R. L. *et al.* Isolate-specific S-antigen of *Plasmodium falciparum* contains a repeated sequence of eleven amino acids. *Nature* **306**, 751–756 (1983).
- Taylor, H. M. *et al.* *Plasmodium falciparum* homologue of the genes for *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins, which is transcribed but not translated. *Infect. Immun.* **69**, 3635–3645 (2001).
- Kaneko, O. *et al.* The high molecular mass rhoptry protein, RhopH1, is encoded by members of the clag multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **118**, 223–231 (2001).
- Trenholme, K. R. *et al.* clag9: A cytoadherence gene in *Plasmodium falciparum* essential for binding of parasitized erythrocytes to CD36. *Proc. Natl Acad. Sci. USA* **97**, 4029–4033 (2000).
- Klembs, M. & Goldberg, D. E. Biological roles of proteases in parasitic protozoa. *Annu. Rev. Biochem.* **71**, 275–305 (2002).
- Banerjee, R. *et al.* Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl Acad. Sci. USA* **99**, 990–995 (2002).
- Rosenthal, P. J., Sijwali, P. S., Singh, A. & Shenai, B. R. Cysteine proteases of malaria parasites: targets for chemotherapy. *Curr. Pharm. Des.* **8**, 1659–1672 (2002).
- Eggleston, K. K., Duffin, K. L. & Goldberg, D. E. Identification and characterization of falcylisin, a metalloprotease involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **274**, 32411–32417 (1999).
- Sinden, R. E., Butcher, G. A., Billker, O. & Fleck, S. L. Regulation of infectivity of *Plasmodium* to the mosquito vector. *Adv. Parasitol.* **38**, 53–117 (1996).
- Billker, O., Shaw, M. K., Margo, G. & Sinden, R. E. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. *Nature* **392**, 289–292 (1998).
- Krungkrai, J., Prapunwattana, P. & Krungkrai, S. R. Ultrastructure and function of mitochondria in gametocytic stage of *Plasmodium falciparum*. *Parasite* **7**, 19–26 (2000).
- Kapke, S. H. *et al.* Exploring the transcriptome of the malaria sporozoite stage. *Proc. Natl Acad. Sci. USA* **98**, 9895–9900 (2001).
- Dessens, J. T. *et al.* CTRP is essential for mosquito infection by malaria ookinetes. *EMBO J.* **18**, 6221–6227 (1999).
- Deitsch, K. W. & Welles, T. E. Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **76**, 1–10 (1996).
- Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 9333–9338 (1999).
- Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome

- expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26**, 183–186 (2000).
- Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289–1292 (2001).
- Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **31**, 180–183 (2002).
- Hernandez-Rivas, R. *et al.* Expressed var genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol. Cell Biol.* **17**, 604–611 (1997).
- del Portillo, H. A. *et al.* A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842 (2001).
- Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
- Kanaani, J. & Ginsburg, H. Metabolic interconnection between the human malarial parasite *Plasmodium falciparum* and its host erythrocyte. *J. Biol. Chem.* **264**, 3194–3199 (1989).
- Decherer, K. J. *et al.* Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell Biol.* **19**, 967–978 (1999).
- Lockhart, D. J. & Winzler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).
- Pacheco, N. D., Strome, C. P., Mitchell, F., Bawden, M. P. & Beaudoin, R. L. Rapid, large-scale isolation of *Plasmodium berghei* sporozoites from infected mosquitoes. *J. Parasitol.* **65**, 414–417 (1979).
- Mellouk, S. *et al.* Evaluation of an *in vitro* assay aimed at measuring protective antibodies against sporozoites. *Bull. World Health Organ.* **68** Suppl., 52–59 (1990).
- Rabilloud, T. *et al.* Analysis of membrane proteins by two-dimensional electrophoresis: comparison of the proteins extracted from normal or *Plasmodium falciparum*-infected erythrocyte ghosts. *Electrophoresis* **20**, 3603–3610 (1999).
- Blackman, M. J. Purification of *Plasmodium falciparum* merozoites for analysis of the processing of merozoite surface protein-1. *Methods Cell Biol.* **45**, 213–220 (1994).
- Haynes, J. D. & Moch, J. K. Automated synchronization of *Plasmodium falciparum* parasites by culture in a temperature-cycling incubator. *Methods Mol. Med.* **72**, 489–497 (2002).
- Haynes, J. D., Moch, J. K. & Smoot, D. S. Erythrocytic malaria growth or invasion inhibition assays with emphasis on suspension culture GIA. *Methods Mol. Med.* **72**, 535–554 (2002).
- Carter, R., Ranford-Cartwright, L. & Alano, P. The culture and preparation of gametocytes of *Plasmodium falciparum* for immunochemical, molecular, and mosquito infectivity studies. *Methods Mol. Biol.* **21**, 67–88 (1993).
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
- Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
- Eisenhaber, B., Bork, P. & Eisenhaber, F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng.* **11**, 1155–1161 (1998).
- Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We are grateful to J. Graumann, R. Sadygov, G. Chukkappalli, A. Majumdar and R. Sinkovits for computer programming; C. Deciu for the probability calculations; and C. Delahunty and C. Vieille for critical reading of the manuscript. The authors acknowledge the support of the Office of Naval Research, the US Army Medical Research and Materiel Command, and the National Institutes of Health (to J.R.Y.). J.D.R. is funded by a Wellcome Trust Prize Studentship. We thank the scientists and funding agencies comprising the international Malaria Genome Project for making sequence data from the genome of *P. falciparum* clone 3D7 public before publication of the completed sequence. The opinions expressed are those of the authors and do not reflect the official policy of the Department of the Navy, Department of Defense, or the US government.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.Y. (e-mail: jjyates@scripps.edu).

Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13

N. Hall*, A. Pain*, M. Berriman*, C. Churcher*, B. Harris*, D. Harris*, K. Mungall*, S. Bowman*†, R. Atkin*, S. Baker*, A. Barron*, K. Brooks*, C. O. Buckee*, C. Burrows*, I. Cherevach*, C. Chillingworth*, T. Chillingworth*, Z. Christodoulou‡, L. Clark*, R. Clark*, C. Corton*, A. Cronin*, R. Davies*, P. Davis*, P. Dear§, F. Dearden*, J. Doggett*, T. Feltwell*, A. Goble*, I. Goodhead*, R. Gwilliam*, N. Hamlin*, Z. Hance*, D. Harper*, H. Hauser*, T. Hornsby*, S. Holroyd*, P. Horrocks‡, S. Humphray*, K. Jagels*, K. D. James*, D. Johnson*, A. Kerhornou*, A. Knights*, B. Konfortov§, S. Kyes‡, N. Larke*, D. Lawson*, N. Lennard*, A. Line*, M. Maddison*, J. McLean*, P. Mooney*, S. Moule*, L. Murphy*, K. Oliver*, D. Ormond*, C. Price*, M. A. Quail*, E. Rabbinowitsch*, M.-A. Rajandream*, S. Rutter*, K. M. Rutherford*, M. Sanders*, M. Simmonds*, K. Seeger*, S. Sharp*, R. Smith*, R. Squares*, S. Squares*, K. Stevens*, K. Taylor*, A. Tivey*, L. Unwin*, S. Whitehead*, J. Woodward*, J. E. Sulston*, A. Craig||‡, C. Newbold‡ & B. G. Barrell*

* The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

‡ The Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

§ MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK
|| Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Since the sequencing of the first two chromosomes of the malaria parasite, *Plasmodium falciparum*^{1,2}, there has been a concerted effort to sequence and assemble the entire genome of this organism. Here we report the sequence of chromosomes 1, 3–9 and 13 of *P. falciparum* clone 3D7—these chromosomes account for approximately 55% of the total genome. We describe the methods used to map, sequence and annotate these chromosomes. By comparing our assemblies with the optical map, we indicate the completeness of the resulting sequence. During annotation, we assign Gene Ontology terms to the predicted gene products, and observe clustering of some malaria-specific terms to specific chromosomes. We identify a highly conserved sequence element found in the intergenic region of internal *var* genes that is not associated with their telomeric counterparts.

Contiguous DNA sequences (contigs) have been obtained for chromosomes 1, 3, 4, 5 and 9, whereas chromosomes 6, 7, 8 and 13 contain a few gaps; most contigs have been ordered and oriented. Table 1 shows the status and content of the chromosomes at the time of writing. As we were unable to produce unbroken sequence from telomere to telomere for all nine chromosomes, contiguous ‘pseudo-chromosomes’ were constructed by artificially joining all contigs that could be mapped to an individual chromosome. In most cases, the order and orientation of the contigs could be inferred using mapping data^{3–5} or read-pair information. Small contigs (of less than 5 kilobases, kb) that could not be mapped onto a chromosome have not been included in the analysis, and thus a small number of genes on the unmapped contigs will be missing from the genome sequence. The construction of pseudo-chromosomes does, however, have the advantage of allowing a global analysis of chromosome structure, and also removes redundancy from the analysis that would otherwise occur owing to contamination between chromosomes during purification and aberrant contigs formed during assembly.

A comparison of the optical maps for the finished chromosomes with virtual restriction digests with two enzymes of the assembled sequences show good agreement (Fig. 1). A misassembly in chromosome 4 is apparent from both comparisons, which we have localized to a region in an internal *var* gene repeat. The

depth of coverage in this area suggests that there is a 50-kb perfect repeat. Chromosome 9 has a deletion of 100 kb in comparison with the *Bam*HI optical map, but it compares well with the *Nhe*I map, and with the sequence tagged site (STS) markers and the yeast artificial chromosome (YAC) map. The data strongly suggest that this anomaly is due to an optical mapping error, rather than a problem with the chromosome sequence.

The sizes of the pseudo-chromosomes 6, 7 and 8 also compare well with the predictions from the optical map. Chromosome 13 is 400 kb smaller than the predicted size in the *Nhe*I map, but only 10 kb smaller than the predicted size from the *Bam*HI map. Thus size comparisons between optical maps and digests reveal that very few data are missing from the chromosome assemblies (Fig. 1). When comparing contig order and orientation with the optical map of unfinished chromosomes, many more outliers are visible on the scatter plots (Fig. 1 and Table 1). Only chromosomes 13 and 6 have r^2 values of less than 0.8 in correlation analysis, both against the *Bam*HI maps. Thus for the most part, the contigs are ordered and oriented correctly.

Chromosomes 6, 7 and 8 do not resolve on pulsed field gel electrophoresis, and therefore they were sequenced as a group. Because of this we were unable to group contigs sufficiently to initiate gap closure. In order to overcome this problem, a HAPPY map^{6–8} was created, using data from the genome sequence to design primers. (HAPPY mapping allows the order and spacing of STS markers to be determined accurately, by following their segregation among roughly haploid samples of randomly fragmented DNA, using the polymerase chain reaction.) In the first round of mapping, 496 probes were generated which could be arranged on 61 linkage groups with 343 singletons at a lod (log of odds) threshold of 4. A further 30 probes were incorporated to increase the number of linkage groups to 62 at a lod threshold of 5 with 361 singletons. The large number of singletons produced was due to the high level of extra-chromosomal contamination of the purified chromosomes, which we estimated to be around 40%. Despite this, generation of a HAPPY map for chromosomes 6, 7 and 8 has been an invaluable step in grouping contigs to direct the finishing process.

Although gene predictions and annotations were performed by three different groups as part of the sequencing consortium, the predicted overall protein-coding content of each chromosome was very similar (Table 1). Small differences in coding percentage were seen in part due to chromosome size and thus their respective contributions of the telomeric sequences. The gene structures predicted from each group, assessed by comparing gene size, exon size and intron size, were also largely the same (Table 1). As the sequence for some chromosomes is incomplete, it is possible that exons that overlap gaps may be missed. In some cases where frame-shifts occur within exons, particular effort has been made to check that these are pseudogenes and not caused by sequencing errors. The consistency of annotations across all chromosomes suggests that the quality of sequence has not seriously affected gene identification. We expect the accuracy of sequence of all chromosomes to be very high owing to the depth of read coverage (Table 1). Chromosome maps showing the location and structure of genes along each chromosome are available (Supplementary Information).

Gene Ontology (GO) was used to classify genes across the entire genome, and as GO had not been previously applied for annotating an intracellular parasite, new parasite-specific GO terms were created⁹. The proportion of genes associated with parasite-specific processes or localized in parasite-specific compartments varies between chromosomes (Fig. 2). Whereas most ‘housekeeping’ genes appear to be evenly distributed across the chromosomes (Fig. 2a), chromosome 5 appears to have the highest proportion of genes annotated with apicoplast localization (Fig. 2b). Conversely, and unlike chromosome 4, it has a very low proportion of genes associated with host cell invasion or adhesion (Fig. 2b, c). The

† Present address: Syngenta, Jealott's Hill International Research Centre, Bracknell RG42 6EY, UK.

uneven distribution of apicoplast targeted genes on chromosome 5 involves non-orthologous genes, whereas the clustering of genes involved in host cell invasion or adhesion results from duplications of gene families such as *variant antigen (var)* and *repetitive interspersed family (rif)* genes.

We have identified two previously undescribed clustered gene families; one on chromosome 9 and one on chromosome 13. On chromosome 9, there are 7 copies of a putative protein kinase which show 25–46% amino-acid identity to each other; four of these genes have a predicted signal peptide. Proteomic analysis has shown expression of two of these genes (PFI0105c and PFI0135c)¹⁰. Chromosome 13 contains a tandem array of 5 paralogous genes including *msp7* (ref. 11) with 15–30% identity to each other. Expression of one of these MSP7-like proteins (MAL13P1.174) has been detected, by proteomic studies, during the asexual stage¹².

The significance of the physical localization and function of these different genes is unknown, so further studies of their expression pattern and cellular localization are required. Protein alignments of these families are available (Supplementary Information).

Bowman *et al.*² deduced a consensus pattern of repeats and coding regions for the subtelomeric regions of chromosomes 2 and 3. The overall arrangement of *var*, *rif* and *subtelomeric variable open reading frame (stevor)* genes is conserved in nearly all telomeres, but the number and orientation of gene families vary. For example, many subtelomeres contain multiple *var* genes, and some have inverted *var* genes. The right-hand telomere of chromosome 5 has a truncated telomere with a partial inverted *var* gene adjacent to the telomeric repeat, with no rep11 or rep20 repeat units. The telomere-associated repeat elements are involved in co-localization of telomeres within the nucleus^{13,14}. This may aid chromosome

Table 1 **Summary statistics**

	Value									
	Whole genome	Chr. 1	Chr. 3	Chr. 4	Chr. 5	Chr. 6	Chr. 7	Chr. 8	Chr. 9	Chr. 13
The genome										
Size (bp)	22,853,764	643,292	1,060,087	1,204,112	1,343,552	1,377,956	1,350,452	1,323,195	1,541,723	2,747,327
No. of gaps	93	0	0	0	0	8	14	24	0	37
Coverage*	14.5	13.3	10.9	16.8	15.1	16.8	15.8	16.2	17.9	17.2
Mapped YACs	—	15	19	18	16	16	17	23	14	29
HAPPY map linkage groups	—	—	—	—	—	17	7	8	—	—
BamHI map length	—	667.9	1,146.6	1,136.8	1,306.8	1,443.8	1,503.7	1,372.8	1,687.9	2,734.9
<i>r</i> ² BamHI	—	0.994	0.999	0.778	0.998	0.796	0.878	0.986	0.958	0.741
NheI optical map length (kb)	—	683.8	1,083.5	1,311.1	1,394.8	1,494.7	1,493.5	1,331.4	1,600.0	3,171.8
<i>r</i> ² NheI	—	0.999	0.997	0.983	0.998	0.908	0.989	0.878	0.909	0.821
(G + C) content (%)	19.4	20.5	19.9	20.7	19.3	19.7	20.0	19.7	19.0	19.2
No. of genes	5,268	143	239	237	312	312	277	295	365	672
Mean gene length (bp)	2,283.3	1,965.0	2,319.5	2,643.9	2,307.0	2,403.6	2,755.1	2,376.3	2,092.2	2,254.5
Gene density (kb per gene)	4,338.2	4,498.5	4,435.5	5,080.6	4,306.3	4,416.5	4,875.3	4,485.4	4,223.9	4,088.3
Percent coding†	52.6	43.7	52.3	52.0	53.6	54.4	56.5	53.0	49.5	55.1
Genes with introns (%)	53.9	69.9	59.0	58.6	52.6	52.9	56.0	57.3	59.2	52.7
Genes with ESTs (%)	47.4	37.8	51.5	45.1	51.0	52.2	45.5	48.1	52.9	54.6
Gene products detected by proteomics‡ (%)	48.2	50.3	53.1	50.6	54.8	52.8	51.6	55.6	53.4	53.4
Exons										
Number	12,674	373	638	576	736	809	651	784	925	1,656
Mean no. per gene	2.4	2.6	2.7	2.4	2.4	2.6	2.4	2.7	2.5	2.5
(G + C) content (%)	23.7	25.3	23.8	25.2	23.6	23.7	24.1	23.9	23.6	23.1
Mean length (bp)	949.1	753.3	868.9	1,087.9	978.0	927.0	1,172.3	894.2	825.6	914.9
Total length (bp)	12,028,350	280,998	554,355	626,607	719,781	749,937	763,167	701,019	763,644	1,515,033
Introns										
Number	7,406	230	399	339	424	497	374	489	560	984
(G + C) content (%)	13.5	13.5	13.4	13.5	13.6	13.8	13.5	13.6	13.4	13.4
Mean length (bp)	178.7	170.4	163.6	186.3	167.7	169.6	180.9	167.8	172.4	158.1
Total length (bp)	1,323,509	39,183	65,279	63,169	71,122	84,283	67,669	82,031	96,547	155,553
Intergenic regions										
(G + C) content (%)	13.6	14.2	13.6	14.0	13.5	13.9	13.8	13.8	13.2	13.4
Mean length (bp)	1,693.9	1,883.4	1,608.9	1,949.4	1,662.6	1,640.4	1,773.2	1,703.1	1,716.8	1,499.2
RNAs										
No. of tRNA genes	43	0	2	5	5	3	7	0	0	5
No. of 5S rRNA genes	3	0	0	0	0	0	0	0	0	0
No. of 5.8S, 18S, 28S rRNA units	7	1	0	0	1	0	1	2	0	1
The proteome										
Total predicted proteins	5,268	143	239	237	312	312	277	295	365	672
Hypothetical proteins§	3,208	80	140	138	175	168	159	189	219	396
InterPro matches	2,650	64	147	141	151	164	112	147	176	227
Pfam matches	1,746	52	100	96	131	131	91	115	139	ND
Gene Ontology										
Process	1,301	41	58	78	62	77	84	62	83	184
Function	1,244	29	59	60	76	67	66	66	88	189
Component	2,412	88	119	121	140	125	149	145	169	281
Targeted to apicoplast	551	14	29	20	49	17	30	33	43	69
Targeted to mitochondrion	246	3	9	3	20	23	16	17	19	31
Structural features										
Transmembrane domain(s)	1,631	74	79	82	89	92	96	104	117	179
Signal peptide	544	21	33	30	32	31	33	20	46	65
Signal anchor	367	18	9	18	23	23	16	16	34	44

ND, not determined; EST, expressed sequence tag. The optical map lengths were calculated by adding together the lengths of restriction fragments in order to estimate the amount of data missing from each of the unfinished chromosomes. The Pearson's product moment coefficient (r^2) was calculated for each chromosome against each of the optical maps using regression analysis (see Fig. 1). Specialized searches used the following programs and databases: InterPro²⁰; Pfam³⁰; Gene Ontology²⁸. Predictions of apicoplast and mitochondrial targeting were performed using TargetP³¹ and MitoProfil³²; transmembrane domains, TMHMM³³; and signal peptides and signal anchors, SignalP-2.0 (ref. 27).

* Average number of sequence reads per nucleotide.

† Excluding introns.

‡ Percentage of proteins detected in parasite extracts by two independent proteomic analyses^{10,12}.

§ Hypothetical proteins are proteins with insufficient similarity to characterized proteins in other organisms to justify provision of functional assignments.

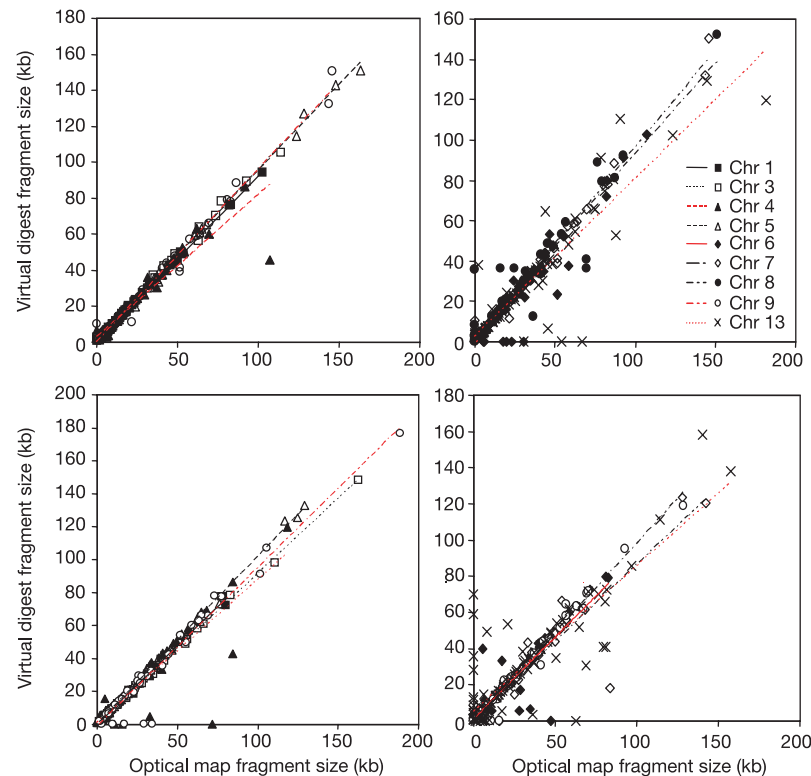


Figure 1 Scatter graphs of virtual restriction digests of completed chromosomes and pseudo-chromosomes against optical map fragment sizes. Top row: completed chromosomes (left) and unfinished chromosomes (right) compared with *NheI* optical map.

Bottom row; as top row but compared with *BamHI* optical map. Each point on the graph represents a restriction fragment compared to its corresponding optical map fragment. The lines show the regression for each chromosome.

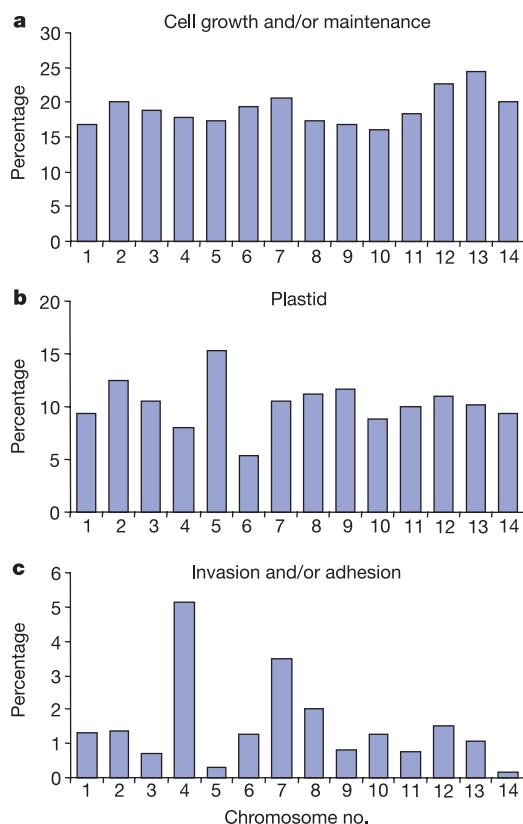


Figure 2 Comparison of the percentage of annotations with specific Gene Ontology terms on each chromosome. **a**, Annotations to 'cell growth and/or maintenance'; **b**, annotations to 'plastid'; **c**, annotations to 'invasion' and/or adhesion.

segregation and increased recombination between subtelomeric genes. Telomere repeats extending from truncated genes are frequently observed in other clones of *P. falciparum*, often leading to transcription of the telomere¹³. This observation suggests that telomere transcription may be involved in telomere maintenance at truncated chromosome ends. As the *var* gene on the right-hand end of chromosome 5 is inverted, there could be transcription of the telomeric repeat.

A putative centromere structure has been predicted in chromosomes 2 and 3 (ref. 2) which is characterized by a 2.6-kb region of 97.3% (A + T) content residing in a gap between coding sequences of at least 9 kb. On inspection of all of the completed chromosomes, we have identified similar structures representing the putative centromeres. There is only ever one per chromosome. All have a region of very high (A + T) content, and a core region of slightly higher (G + C) content, all lying in a gap between coding regions of between 8 and 11 kb. A similar structure has now been identified in the intracellular parasite *Encephalitozoon cuniculi*¹⁵. The discovery of these elements in all contiguous chromosomes, and now in another organism, suggests they have an important role in chromosome maintenance.

Three of the nine chromosomes that were sequenced by us (namely 4, 7 and 8) contain internal arrays of *var* genes. In the intergenic regions of the internal *var* arrays, we have identified a highly conserved, (G + C)-rich (~40% (G + C) content), sequence element of length ~202 bp (Fig. 3). We have also identified three such (G + C)-rich conserved elements on chromosome 12, sequenced in ref. 16 (not shown in Fig. 3). There are in total 15 of these (G + C)-rich elements in the entire *P. falciparum* genome, with not more than one element present in every internal *var* intergenic region. These (G + C)-rich elements are strictly associated with internal *var* arrays, and were not found in subtelomeric *var* genes, nor near the single internal *var* genes on chromosomes 6

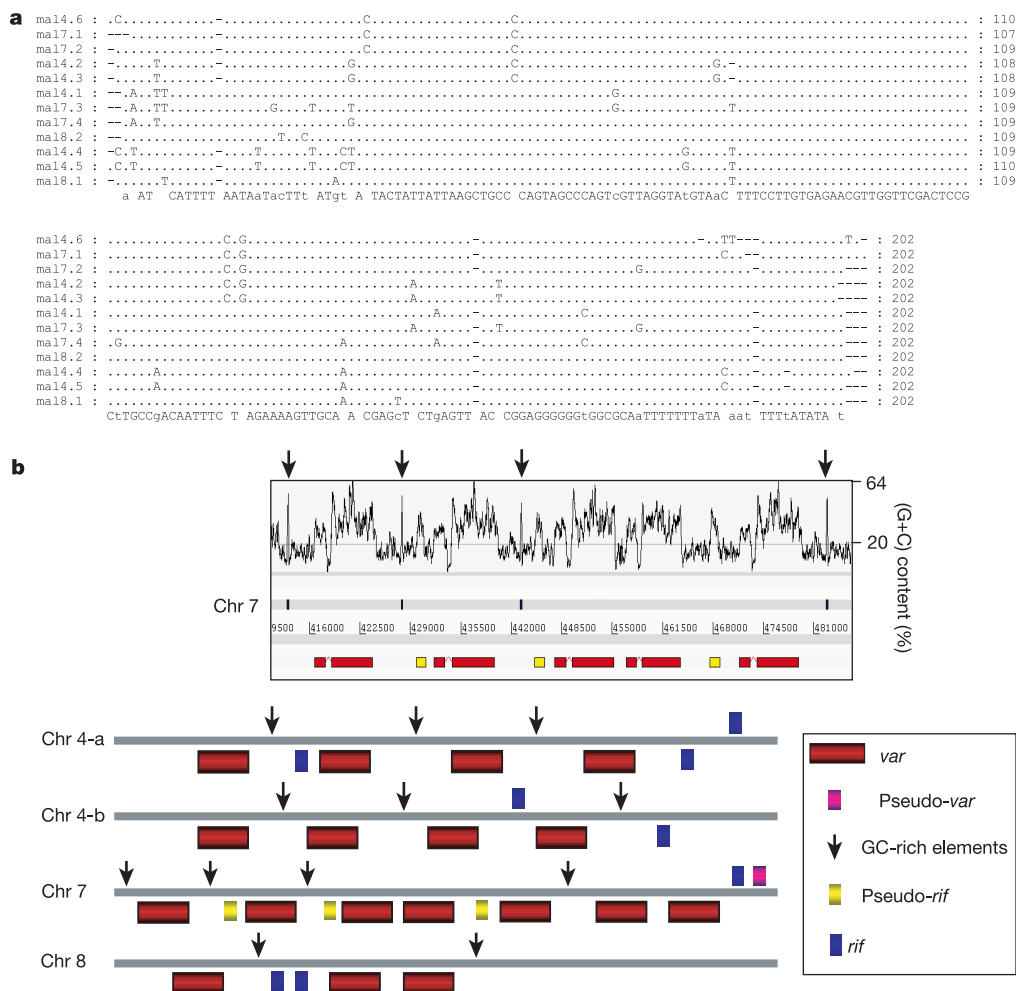


Figure 3 Position and structure of *var*-related (G + C)-rich elements. **a**, Multiple alignment of the (G + C)-rich conserved sequence elements on chromosomes 4, 7 and 8 of *P. falciparum*, using CLUSTAL. Only the non-identical nucleotides across all 12 (G + C)-rich conserved sequence elements are indicated in the alignment, with the consensus sequence indicated at the bottom. The upper-case letters in the consensus sequence denote complete identity across all the (G + C)-rich elements presented in the alignment. Each of these sequence elements is represented with a unique identifier, representing its specific origin. **b**, Location of the (G + C)-rich conserved sequence elements in the intergenic region of internal *var* gene clusters on chromosomes 4, 7 and 8

and 12. There is no obvious systematic order of the location of these (G + C)-rich sequence elements with respect to adjacent *var* genes in terms of proximity or direction of transcription of the *var* genes. The specific positioning of these conserved sequence elements between internal *var* genes suggests a possible regulatory function, although a standard BLASTN query in public databases showed no significant similarity to previously identified RNA genes or gene regulatory elements. The (G + C)-rich element does have the potential to form secondary structures when analysed using the MFOLD program (<http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html>) (data not shown). This could indicate that the (G + C)-rich element is a hitherto unknown transcribed RNA species. *Cis*-acting (G + C)-rich gene regulatory elements have been shown to function as important transcriptional regulators present in the promoter, enhancer and locus control regions of many eukaryotic genes from several species (see ref. 17 for a review). The interaction between specific sites along a DNA molecule has been shown to have a crucial role in the regulation of genetic

of *P. falciparum*. Top panel, four (G + C)-rich sequence elements in the intergenic regions of internal *var* gene cluster on chromosome 7. The arrowheads indicate the peaks in the (G + C) plot, corresponding to the location of the (G + C)-rich conserved sequence elements. The exact location of the neighbouring *var* and pseudo-*rif* genes are marked with red and yellow boxes, respectively. Bottom panel, a schematic diagram representing the relative positions of the internal *var* and *rif* genes and the conserved (G + C)-rich sequence elements on chromosomes 4, 7 and 8 (not to scale). The *var* or *rif* genes are placed either on top or bottom of the grey bars, depending on the direction of transcription.

processes such as DNA replication, site-specific recombination and transposition in other organisms¹⁸. Control of gene expression through DNA loop formation has also been shown in other organisms¹⁸, while in *P. falciparum* regulation of *var* gene expression by cooperative gene silencing elements in *var* gene introns¹⁹, or by a 5' flanking *var* gene region regulatory element, has also been described²⁰. The potential of the (G + C)-rich sequences to form DNA secondary structures supports a possible function as regulatory elements in *var*-related genetic processes in *P. falciparum*. □

Methods

Sequencing

The DNA was cloned and sequenced according to methods described elsewhere^{2,21}. Derived contigs were ordered according to previously derived genetic, optical and physical maps³⁻⁵. For all unfinished chromosomes, assemblies were screened against mapped contigs to remove extra-chromosomal contamination. For chromosomes 6, 7 and 8 a HAPPY map was generated to assist ordering; briefly, agarose-embedded genomic DNA was released by melting at 65 °C, sheared gently into fragments with a mean size of ~50 kb, and 88 samples, each containing ~0.7 genome-equivalents of fragments, were taken (a

further 8 samples were DNA-free controls). These samples (the mapping panel) were preamplified by PEP (primer extension preamplification), diluted and dispensed into 30 replica panels. Each replica was screened for between 50 and 100 markers using a two-phase polymerase chain reaction (multiplexed forward and reverse primers in phase 1, followed by dilution and a second phase for one marker at a time, using an internal forward primer and the reverse primer). Pairwise lod scores between markers were calculated, linkage groups identified, and maps of each group of three or more markers computed, essentially as described previously^{7,8}.

Annotation

Genome annotation was carried out using Artemis²². Genes were identified by manual curation of the output of the software packages Genefinder (P. Green, unpublished work), GlimmerM²³ and phat²⁴. Functional assignments were based on assessment of BLAST and FASTA searches against public databases and domain predictions using InterProScan²⁵, TMHMM²⁶ and SignalP²⁷.

Gene Ontology (GO) terms²⁸ were manually assigned to gene products for all 14 chromosomes. First, candidate GO terms were selected by sequence-similarity searching a database of peptide sequences and their previously assigned GO terms, drawn from the following databases: Flybase, Mouse Genome Informatics, *Saccharomyces* Genome Database, Swissprot and The *Arabidopsis* Information Resource. After visual inspection of sequence alignments, suitable terms were either assigned directly from the candidate list, or alternatively, higher or lower granularity terms were selected directly from the ontology. When previously characterized genes were identified, terms were selected as above, but alternative experimental evidence codes were used to reflect the fact that the inferences were no longer based on sequence similarity. Some GO terms were also assigned automatically. In particular, 'membrane' was assigned using the transmembrane helix prediction tool TMHMM 2.0 (ref. 26).

Received 31 July; accepted 2 September 2002; doi:10.1038/nature01095.

- Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
- Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
- Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
- Lai, Z. *et al.* A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature Genet.* **23**, 309–313 (1999).
- de Bruin, D., Lanzer, M. & Ravetch, J. V. Characterization of yeast artificial chromosomes from *Plasmodium falciparum*: construction of a stable, representative library and cloning of telomeric DNA fragments. *Genomics* **14**, 332–339 (1992).
- Glockner, G. *et al.* Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* **418**, 79–85 (2002).
- Piper, M. B., Bankier, A. T. & Dear, P. H. A HAPPY map of *Cryptosporidium parvum*. *Genome Res.* **8**, 1299–1307 (1998).
- Konfortov, B. A., Cohen, H. M., Bankier, A. T. & Dear, P. H. A high-resolution HAPPY map of *Dictyostelium discoideum* chromosome 6. *Genome Res.* **10**, 1737–1742 (2000).
- Berriman, M., Aslett, M. & Ivens, A. Parasites are GO. *Trends Parasitol.* **17**, 463–464 (2001).
- Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
- Pacheb, J. A. *et al.* The 22 kDa component of the protein complex on the surface of *Plasmodium falciparum* merozoites is derived from a larger precursor, merozoite surface protein 7. *Mol. Biochem. Parasitol.* **117**, 83–89 (2001).
- Lasorder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high accuracy mass spectrometry. *Nature* **419**, 531–542 (2002).
- Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marín, J. C. & Scherf, A. A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J.* **21**, 815–824 (2002).
- O'Donnell, R. A. *et al.* A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes. *EMBO J.* **21**, 1231–1239 (2002).
- Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
- Hyman, R., Fung, E. & Dennis, R. W. *et al.* Sequence of *Plasmodium falciparum* chromosome 12. *Nature* **419**, 534–536 (2002).
- Hapgood, J. P., Riedemann, J. & Scherf, S. D. Regulation of gene expression by GC-rich DNA cis-elements. *Cell Biol. Int.* **25**, 17–31 (2001).
- Adhya, S. Multipartite genetic control elements: communication by DNA loop. *Annu. Rev. Genet.* **23**, 227–2250 (1989).
- Deitsch, K. W., Calderwood, M. S. & Wellems, T. E. Malaria. Cooperative silencing elements in var genes. *Nature* **412**, 875–876 (2001).
- Vazquez-Macias, A. *et al.* A distinct 5' flanking var gene region regulates *Plasmodium falciparum* variant erythrocyte surface antigen expression in placental malaria. *Mol. Microbiol.* **45**, 155–167 (2002).
- Quail, M. A. M13 cloning of mung bean nuclease digested PCR fragments as a means of gap closure within A/T-rich, genome sequencing projects. *DNA Seq.* **12**, 355–359 (2001).
- Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
- Salzberg, S. L., Pertea, M., Delcher, A. L., Gardner, M. J. & Tettelin, H. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59**, 24–31 (1999).
- Cawley, S. E., Wirth, A. I. & Speed, T. P. Phat—a gene finding program for *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **118**, 167–174 (2001).
- Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
- Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).
- Nielsen, H., Brunak, S. & von Heijne, G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**, 3–9 (1999).

- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
- Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
- Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
- Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
- Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank the staff in the computer support and software development groups; J. Thompson and A. Cowman for gifts of YAC clones and for advice; D. Schwartz for optical map data; X. Su for genetic map information; Y. Shaw for help with Fig. 1; M. Harris and M. Ashburner for assistance with the parasite specific GO terms; O. White and M. Gardner for Table 1 and supplementary figures; the other members of the Malaria Genome Sequencing Consortium for discussions; and The Wellcome Trust Plasmodium Genome Mapping Consortium. This work was supported by the Wellcome Trust.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.H. (e-mail: nh1@sanger.ac.uk). Sequences have been deposited in the EMBL database with accession numbers AL844501 (chromosome 1), AL844502 (chromosome 3), AL844503 (chromosome 4), AL844504 (chromosome 5), AL844505 (chromosome 6), AL844506 (chromosome 7), AL844507 (chromosome 8), AL844508 (chromosome 9) and AL844509 (chromosome 13). Other information is available at http://www.sanger.ac.uk/Projects/P_falciparum.

Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14

Malcolm J. Gardner*, Shamira J. Shallom*, Jane M. Carlton*, Steven L. Salzberg*, Vishvanath Nene*, Azadeh Shoaibi*, Anne Ciecko*, Jeffery Lynn*, Michael Rizzo*, Bruce Weaver*, Behnam Jarrahi*, Michael Brenner*, Babak Parvizi*, Luke Tallon*, Azita Moazzez*, David Granger*, Claire Fujii*, Cheryl Hansen*, James Pederson†, Tamara Feldblyum*, Jeremy Peterson*, Bernard Suh*, Sam Angiuoli*, Mihaela Pertea*, Jonathan Allen*, Jeremy Selengut*, Owen White*, Leda M. Cummings‡, Hamilton O. Smith‡, Mark D. Adams‡, J. Craig Venter‡, Daniel J. Carucci†, Stephen L. Hoffman† & Claire M. Fraser*

* The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

† Malaria Program, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, Maryland 20910-7500, USA

The mosquito-borne malaria parasite *Plasmodium falciparum* kills an estimated 0.7–2.7 million people every year, primarily children in sub-Saharan Africa. Without effective interventions, a variety of factors—including the spread of parasites resistant to antimalarial drugs and the increasing insecticide resistance of mosquitoes—may cause the number of malaria cases to double over the next two decades¹. To stimulate basic research and facilitate the development of new drugs and vaccines, the genome of *Plasmodium falciparum* clone 3D7 has been sequenced using a chromosome-by-chromosome shotgun strategy^{2–4}. We report

‡ Present addresses: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA (L.M.C.); Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850, USA (H.O.S., M.D.A.); The Center for the Advancement of Genomics, 1901 Research Boulevard, 6th Floor, Rockville, Maryland 20850, USA (J.C.V.); Sanaria, 308 Argosy Drive, Gaithersburg, Maryland 20878, USA (S.L.H.).

here the nucleotide sequences of chromosomes 10, 11 and 14, and a re-analysis of the chromosome 2 sequence⁵. These chromosomes represent about 35% of the 23-megabase *P. falciparum* genome.

P. falciparum chromosomes were resolved on preparative pulsed field gels, and used to prepare shotgun libraries of 1–2-kilobase (kb) DNA fragments in plasmid vectors. Sequences of randomly selected clones were assembled, and gaps were closed using primer walking on plasmid templates or polymerase chain reaction (PCR) products. The cross-contamination of the chromosomal libraries with sequences from other chromosomes (up to 25%) and the high (A + T) content (80.6%) of *P. falciparum* DNA caused extreme difficulties in the gap closure process. Intergenic regions and introns frequently contained long runs of up to 50 consecutive A or T residues that were difficult to clone and sequence. The high (A + T) content of the chromosomes also prevented the construction of large insert libraries that could be used to construct scaffolds of ordered and oriented contiguous DNA sequences (contigs) during assembly. Similar but more severe problems were reported in the sequencing of the (A + T)-rich chromosome 2 of the slime mould *Dictyostelium discoideum*⁶, illustrating the need to develop better

methods for the cloning and sequencing of very (A + T)-rich genomes. The reported sequences contain three or four short gaps (<2 kb) in each chromosome. Contigs comprising these chromosomes were joined end-to-end before annotation. Efforts to close the remaining gaps will continue.

Examination of the sequences of chromosomes 2, 10, 11 and 14 revealed that the structure of these chromosomes was similar to that of the other chromosomes. All contained the 97–99% (A + T) putative centromeric sequences reported previously⁷. Conserved subtelomeric sequences² were observed in chromosomes 2, 10 and 11, but most of these elements had been deleted from both ends of chromosome 14. The termini of chromosome 14 consisted of telomeric hexamer repeats fused directly to truncated *var* (variant antigen) genes. Deletions of this type are thought to be due to chromosome breakage and healing events that occur during *in vitro* cultivation of the parasite.

Annotation procedures have improved since the publication of the *P. falciparum* chromosome 2 sequence⁵. A gene finding program, phat (pretty handy annotation tool⁸), was developed, supplementing the GlimmerM program⁹ used previously. In this work, GlimmerM and phat were retrained on a larger training set of well-

Table 1 Summary statistics

Feature	Value				
	Whole genome	Chromosome 2	Chromosome 10	Chromosome 11	Chromosome 14
The genome					
Size (bp)	22,853,764	947,102	1,694,445	2,035,250	3,291,006
No. of gaps	93	0	4	3	3
Coverage*	14.5	11.1	15.6	11.3	9.2
(G + C) content (%)	19.4	19.7	19.7	19.0	18.4
No. of genes	5,268	223 (209)	403	492	769
Mean gene length (bp)†	2,283.3	2,079.1 (2,105.1)	2,085.8	2,127.7	2,315.1
Gene density (bp per gene)	4,338.2	4,247.1 (4,531.6)	4,204.6	4,136.7	4,279.6
Percent coding	52.6	49.0 (46.5)	49.6	51.4	54.1
Genes with introns (%)	53.9	57.0 (43.1)	51.4	50.4	49.9
Genes with ESTs (%)	49.1	46.2	48.1	48.4	46.9
Gene products detected by proteomics‡	51.8	43.5	49.1	51.0	52.1
Exons					
Number	12,674	510 (353)	892	1,094	1,757
Mean no. per gene	2.4	2.3 (1.7)	2.2	2.2	2.3
(G + C) content (%)	23.7	24.4 (24.3)	24.5	23.5	22.8
Mean length (bp)	949.1	909.1 (1,246.3)	942.3	956.9	1,013.3
Total length (bp)	12,028,350	463,647 (439,944)	840,576	1,046,814	1,780,305
Introns					
Number	7,406	287 (144)	489	602	988
(G + C) content (%)	13.5	13.4 (13.4)	13.6	13.7	13.5
Mean length (bp)	178.7	202.4 (208.4)	234.5	189.4	185.5
Total length (bp)	1,323,509	58,080 (30,006)	114,676	114,012	183,240
Intergenic regions					
(G + C) content (%)	13.6	13.5 (14.1)	13.6	14.1	13.2
Mean length (bp)	1,693.9	1,702.3 (2,063.2)	1,678.5	1,768.5	1,717.2
RNAs					
No. of tRNA genes	43	1	0	2	2
No. of 5S rRNA genes	3	0	0	0	3
No. of 5.8S, 18S and 28S rRNA units	7	0	0	1	0
The proteome					
Total predicted proteins	5,268	223	403	492	769
Hypothetical proteins§	3,208	121	265	339	485
InterPro matches	2,650	116	210	283	455
Pfam matches	1,746	77	133	184	275
Gene Ontology					
Process	1,301	63	89	110	168
Function	1,244	54	74	95	174
Component	2,412	120	181	220	308
Targeted to apicoplast	551	28	36	52	73
Targeted to mitochondrion	246	10	13	17	33
Structural features					
Transmembrane domain(s)	1,631	87	133	141	202
Signal peptide	544	28	41	52	63
Signal anchor	367	19	32	31	51

Numbers in parentheses under chromosome 2 indicate values obtained in the previous annotation⁵. Specialized searches used the following programs and databases: InterPro²¹, Pfam¹⁹ and Gene Ontology²². Predictions of apicoplast and mitochondrial targeting were performed using TargetP²⁶ and MitoProtII²⁵; transmembrane domains, TMHMM²⁴; and signal peptides and signal anchors, SignalP-2.0 (ref. 23).

*Average number of sequence reads per nucleotide. EST, expressed sequence tag.

†Excluding introns.

‡Percent of proteins detected in parasite extracts by two independent proteomic analyses^{29,30}.

§Hypothetical proteins are proteins with insufficient similarity to characterized proteins in other organisms to justify provision of functional assignments.

characterized genes, complementary DNAs (cDNAs) and products of PCR with reverse transcription (RT-PCR) (total length 540 kb) than was used in the earlier work. A program called Combiner was used to evaluate the GlimmerM and phat predictions, as well as the results of searches against nucleotide and protein databases, to construct consensus gene models. To assess the effect of these modifications, chromosome 2 was re-annotated and the results were compared with the previous annotation.

Application of these automated annotation procedures and manual curation of the resulting gene models for chromosome 2 produced 223 gene models. The revised procedures detected 21 genes not predicted previously, and 13 of the existing chromosome 2 models collapsed into six models in the new annotation. Of the 21 new gene models, all but one had no significant similarity to proteins in a non-redundant amino-acid database. However, at least a portion of each of the 21 gene models had been predicted independently by both GlimmerM and phat, suggesting that many of these models were likely to represent coding sequences. On the other hand, five of the new gene models encoded proteins less than 100 amino acids in length, and may be less likely to encode proteins.

Another major difference was the detection of additional small exons. In the earlier annotation of chromosome 2, the 209 predicted genes contained 353 exons, or an average of 1.7 exons per gene. The revised procedures reported here revealed 510 exons, or 2.3 exons per gene; 60% of the new exons were predicted to be additions to the gene models reported previously. Most cases involved the addition of one or two exons per gene. In three notable cases, however, 7 to 12 small exons were added to the earlier gene models, and almost all of the new exons had been predicted by both of the gene finding programs. Overall, use of the revised annotation procedures resulted in the detection of additional genes and many small exons, which is reflected in the higher gene density and shorter mean exon length in the newly annotated chromosome 2 sequence compared with the previous annotation (Table 1). Despite these improvements in software and training sets, gene finding in *P. falciparum* remains challenging, and the gene structures presented here should be regarded as preliminary until confirmed by sequence information obtained from cDNAs or RT-PCR experiments¹⁰. Accurate prediction of the 5' ends of genes is particularly difficult. Generation of larger training sets, including additional expressed sequence tags (ESTs) and full-length cDNAs, would greatly improve the sensitivity and accuracy of gene predictions.

These annotation procedures were also applied to the analysis of chromosomes 10, 11 and 14 (Table 1; maps of these chromosomes are available as Supplementary Information). The 10 short gaps in the chromosomes should not have interfered with the gene predictions; only the genes adjacent to the gaps might have been affected. All three chromosomes were similar in terms of gene density, coding percentage and other parameters. A complete description of the parasite genome is contained in the accompanying Article².

Annotation of chromosomes 10, 11 and 14 revealed four proteins with sequence similarity to SR proteins, a family of conserved splicing factors that contain RNA-binding domains and a protein interaction domain rich in Ser and Arg residues (SR domain; PF10_0047, PF10_0217, PF11_0200, PF14_0656). Three additional putative SR proteins were identified on chromosomes 5 and 13 (PFE0160c, PFE0865c, MAL13P1.120). SR proteins are thought to bind to exonic splicing enhancers (ESEs), short (6–9 bp) sequences within exons that assist in the recognition of nearby splice sites, and to interact with components of the spliceosome¹¹. ESEs have previously been characterized only in multicellular organisms. To determine whether *P. falciparum* may use ESEs as part of its splicing machinery, a Gibbs sampling algorithm for motif detection¹² was applied to a set of *P. falciparum* exons to detect any exonic splicing enhancers (ESEs). The exons were extracted from the set of well-characterized genes used to train the GlimmerM gene finder.

Regions of 50 bp regions were selected from both ends of the internal exons and divided into two different data sets, representing the exon regions adjacent to both 5' and 3' splice sites. At least 10 runs of the Gibbs sampler were performed for each data set in order to identify the most probable motif with a length of 5–9 nucleotides. The motif with the highest maximum *a posteriori* probability was retained. This analysis identified a motif with the consensus GAAGAA, which is identical to ESEs found in human exons^{13,14}. The identification of several putative SR proteins, and sequences identical to the ESEs in humans, suggests that some features of exon recognition and splicing observed in higher eukaryotes may be conserved in *P. falciparum*. □

Methods

Sequencing and closure

P. falciparum clone 3D7 was selected for sequencing because it can complete all phases of the life cycle, and had been used in a genetic cross¹⁵ and the Wellcome Trust Malaria Genome Mapping Project¹⁶. High-molecular-mass genomic DNA was subjected to electrophoresis on preparative pulsed field gels, and chromosomes were excised. DNA was extracted from the gel, sheared, and cloned into the pUC18 vector as described⁵ (chromosomes 2, 14) or into a modified pUC18 vector via *Bst*XI linkers (chromosomes 10, 11). Sequences were assembled and gaps were closed by primer walking on plasmid DNAs or genomic PCR products, or by transposon insertion⁵. Ordering of contigs was facilitated by the use of sequence tagged sites¹⁶ and microsatellite markers¹⁷. The final assembly of each chromosome was verified by comparison with *Bam*HI and *Nhe*I optical restriction maps¹⁸. The average difference in size between the experimentally determined restriction fragments and the fragments predicted from the sequence was approximately 5–6% for chromosomes 11 and 14 for both enzymes. For chromosome 10, the average difference in fragment sizes was 6.1% for the *Nhe*I map, but the *Bam*HI optical and prediction restriction maps could not be aligned. Because the *Nhe*I optical restriction map agreed with that predicted from the sequence, the chromosome 10 assembly was judged to be correct.

Annotation

GlimmerM⁹ and phat⁸ were trained on 117 *P. falciparum* genes and 39 cDNAs taken from GenBank, plus 32 genes from chromosomes 2 and 3 that had been verified by RT-PCR (provided by R. Huestis and K. Fischer; the training set is available at <http://www.tigr.org/software/glimmerm/data>). The GlimmerM and phat predictions, and sequence alignments of the chromosomes to protein and cDNA databases, were evaluated by the Combiner program. The program used a linear weighting method and dynamic programming to construct consensus gene models that were curated manually using AnnotationStation (AffyMetrix Inc.). Predicted proteins were searched against a non-redundant amino-acid database using BLASTP; other features were identified by searches against the Pfam¹⁹, PROSITE²⁰ and InterPro²¹ databases. The results of all analyses were reviewed using Manatee, a tool that interfaces with a relational database of the information produced by the annotation software. Predicted gene products were manually assigned Gene Ontology²² terms. Signal peptides and signal anchors were predicted with SignalP-2.0 (ref. 23). Transmembrane helices were predicted with TMHMM²⁴. Mitochondrial- and apicoplast-targeted proteins were predicted by MitoProtII²⁵, TargetP²⁶ and PATS²⁷. tRNA-ScanSE²⁸ was used to identify transfer RNAs.

Received 6 August; accepted 2 September 2002; doi:10.1038/nature01094.

1. Breman, J. G. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**, 1–11 (2001).
2. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
3. Hall, N. *et al.* Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature* **419**, 527–531 (2002).
4. Hyman, R. W. *et al.* Sequence of *Plasmodium falciparum* chromosome 12. *Nature* **419**, 534–537 (2002).
5. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
6. Glockner, G. *et al.* Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* **418**, 79–85 (2002).
7. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
8. Carley, S. E., Wirth, A. I. & Speed, T. P. Phat—a gene finding program for *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **118**, 167–174 (2001).
9. Salzberg, S. L., Pertea, M., Delcher, A., Gardner, M. J. & Tettelin, H. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59**, 24–31 (1999).
10. Huestis, R. & Fischer, K. Prediction of many new exons and introns in *Plasmodium falciparum* chromosome 2. *Mol. Biochem. Parasitol.* **118**, 187–199 (2001).
11. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243 (2002).
12. Lawrence, C. E. *et al.* Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**, 208–214 (1993).
13. Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B. & Cooper, T. A. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol. Cell Biol.* **15**, 4898–4907 (1995).
14. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).

15. Walliker, D., Quayk, I., Wellems, T. E. & McCutchan, T. F. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666 (1987).
16. Foster, J. & Thompson, J. The *Plasmodium falciparum* genome project: a resource for researchers. *Parasitol. Today* **11**, 1–4 (1995).
17. Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
18. Lai, Z. *et al.* A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nature Genet.* **23**, 309–313 (1999).
19. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
20. Falquet, L. *et al.* The PROSITE database, its status in 2002. *Nucleic Acids Res.* **30**, 235–238 (2002).
21. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
22. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
23. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
24. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
25. Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
26. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
27. Zuegg, J., Ralph, S., Schmuker, M., McFadden, G. I. & Schneider, G. Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* **280**, 19–26 (2001).
28. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
29. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
30. Lasander, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank our colleagues at The Institute for Genomic Research and the Naval Medical Research Center for support; J. Foster for providing markers for chromosome 14; R. Huestis and K. Fischer for providing RT-PCR data for chromosomes 2 and 3 before publication; and S. Cawley for assistance with phat. This work was supported by the Burroughs Wellcome Fund, the National Institute for Allergy and Infectious Diseases, the Naval Medical Research Center, and the US Army Medical Research and Materiel Command.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.G. (e-mail: gardner@tigr.org). Chromosome sequences have been deposited in GenBank with accession numbers AE001362.2 (chromosome 2), AE014185 (chromosome 10), AE01486 (chromosome 11) and AE01487 (chromosome 14), and in PlasmDB (<http://plasmdb.org>).

Sequence of *Plasmodium falciparum* chromosome 12

Richard W. Hyman, Eula Fung, Aaron Conway, Omar Kurdi, Jennifer Mao, Molly Miranda, Brian Nakao, Don Rowley, Tomoaki Tamaki, Fawn Wang & Ronald W. Davis

Stanford Genome Technology Center, 855 California Avenue, Palo Alto, California 94304 USA, and Departments of Biochemistry and Genetics, Stanford University Medical College, Stanford University, Stanford, California 94305, USA

The human malaria parasite *Plasmodium falciparum* is responsible for the death of more than a million people every year¹. To stimulate basic research on the disease, and to promote the development of effective drugs and vaccines against the parasite, the complete genome of *P. falciparum* clone 3D7 has been sequenced, using a chromosome-by-chromosome shotgun strategy^{2–4}. Here we report the nucleotide sequence of the third largest of the parasite's 14 chromosomes, chromosome 12, which comprises about 10% of the 23-megabase genome. As the most

(A + T)-rich (80.6%) genome sequenced to date, the *P. falciparum* genome presented severe problems during the assembly of primary sequence reads. We discuss the methodology that yielded a finished and fully contiguous sequence for chromosome 12. The biological implications of the sequence data are more thoroughly discussed in an accompanying Article (ref. 3).

At the inception of the Malaria Genome Project, our colleagues at the Institute for Genomic Research (TIGR) and the Wellcome Trust Sanger Institute (WTSI) sequenced *P. falciparum* chromosomes 2 and 3 (refs 5, 6). We chose to sequence the third-largest *P. falciparum* chromosome, chromosome 12, which comprises about 10% of the genome. We made this choice because a 'tiling path' had just been published⁷. (A tiling path is an ordered set of recombinant DNAs covering a large DNA sequence, such as chromosome 12. In this case, the tiling path is composed of yeast artificial chromosomes (YACs) with sequence-tagged sites (STSs, mapped sequence markers).) We predicted that the YACs and the STSs would be helpful in positioning sequence contigs (stretches of contiguous sequence) along *P. falciparum* chromosome 12.

From the published data⁷, we defined a 21 YAC tiling path across *P. falciparum* chromosome 12 (Supplementary Fig. 1). However, we did not want to rely exclusively on sequencing YACs because of three important concerns, which turned out to be warranted. (1) Base changes in the sequence can occur during the construction of any recombinant DNA/YAC, and mutations can occur during passage of any YAC in yeast. (2) One or more YACs in the tiling path might not overlap a neighbouring YAC, creating a physical gap in the sequence. (3) Three of the YACs in the tiling path were derived from *P. falciparum* clone B8 rather than clone 3D7. Polymorphisms between the DNAs of the two strains could hinder the assembly process. Therefore, we devised the following overall strategy. We would sequence random pieces of (that is, use 'shotgun sequencing' on) each of the YACs in the minimum tiling path to low coverage—just enough to establish a 'bin' (a group of related sequences). The bins would give us physical position information across *P. falciparum* chromosome 12. The STSs would give us physical position information within each bin. In addition, we would shotgun-sequence *P. falciparum* chromosome 12 itself. The sequence of each chromosome 12 shotgun sequence 'read' (a sequence of length 100–600 bases derived from a piece of DNA) would be compared to the sequences in each bin. When there was a good match, the read would stay in that bin. This process is highly iterative.

The 21 YACs comprising the minimum tiling path varied considerably in size, with a range of 40–220 kilobases (kb; ref. 7). Our shotgun sequence coverage of the YACs also varied considerably, with a range of 0.5–9.7 YAC coverage (Supplementary Table 1). However, with the exception of four YACs with which we experimented with high coverage early in this project, the shotgun sequence coverage of the remaining YACs was low, as originally planned. In total, there are 14,159 YAC reads (2.6-fold chromosome 12 coverage) supporting the final chromosome 12 sequence. In addition, we produced 69,532 *P. falciparum* chromosome 12 shotgun reads (11.3-fold chromosome 12 coverage) that support the chromosome 12 consensus sequence (Supplementary Table 2). After assembling all of the shotgun sequence data, nearly all of the contigs could be placed unambiguously relative to each other, based on the YAC bins and the STSs. The few remaining contigs were positioned unambiguously by using the genetic map of *P. falciparum* chromosome 12 constructed through the use of microsatellite markers derived from our chromosome 12 sequence^{8,9}. The very few remaining contigs were placed unambiguously by use of the data that accrued during the process of 'finishing' (identifying and replacing all problems in the assembled sequence).

Every part of the assembled sequence of *P. falciparum* chromosome 12 was carefully examined to identify problems in the sequence. These problems were of many types, including (but not limited to) gaps in the sequence, weakly supported sequence,

ambiguities in the sequence, and sequence in only one direction. The problems in the assembled sequence were resolved during the process of finishing. As a result of finishing, we added an additional 7,500 reads (1.09-fold chromosome 12 coverage) in support of the consensus sequence. The final phase of the finishing process was sequence validation. We manually scanned through the *P. falciparum* chromosome 12 consensus sequence, and noted regions (sometimes as large as several hundred base pairs) where we were dissatisfied with the supporting reads. The substantial majority of these regions were composed (almost entirely) of sets of various tandem repeats. As such, there was little reason to try to re-sequence across these regions. However, we were concerned that we may have missed some unique sequence buried in among the repeats.

We therefore undertook a validation process whereby we compared the lengths of PCR (polymerase chain reaction) products with the lengths predicted by the consensus sequence. Manually, we designed three pairs of nested custom primers and performed PCR reactions with those pairs of primers, using total *P. falciparum*

genomic DNA as the template. The lengths of the PCR products were determined experimentally by gel electrophoresis. We could predict the lengths of these PCR products by counting bases in the consensus sequence. Overall, we successfully completed 219 validation reactions. Because we attempted three PCR reactions for every weakly supported place in the consensus sequence, we achieved, at least, one PCR product for virtually all positions. For 201 reactions (92%), the PCR product's measured length was within experimental error of the predicted length. For 18 reactions, representing eight positions on the consensus sequence, the predicted and experimental lengths disagreed by just beyond experimental error. In these cases, we prepared, and sequenced, the appropriate PCR products. For validation using completely independent data, we made use of the two optical restriction enzyme cleavage maps of *P. falciparum* chromosome 12 (ref. 10). (We note that we did not use these optical restriction enzyme cleavage patterns previously: for example, to place contigs relative to each other along chromosome 12.) We normalized the two published cleavage maps to 2.27 megabases, the length of chromosome 12 as determined by our sequence. A comparison of those normalized values to the virtual fragment sizes predicted from our sequence revealed an average discrepancy of less than 6%, which represents excellent agreement.

Our final consensus sequence for *P. falciparum* chromosome 12 is composed of 2,271,477 base pairs (bp) (Table 1). The sequence is completely contiguous; there are no gaps. This sequence is supported by a total of 91,191 reads (14.9-fold chromosome 12 coverage). Overall, the guanine-plus-cytosine (G + C) content of chromosome 12 is 19.3%. As expected from this very low (G + C) content, the *P. falciparum* chromosome 12 sequence contains many long runs of consecutive adenine and thymine residues. Runs of, at least, 20 such bases cover 18% of the chromosome 12 sequence. Bowman *et al.*⁶ were able to identify a region of extremely low (G + C) content as the best candidate location for the centromere of *P. falciparum* chromosome 3. Our chromosome 12 sequence contains an analogous region between base positions 1,282,701 and 1,284,791 (2,090 bp; 0.092% of chromosome 12). That region has a (G + C) content of 1.9%, is composed of the short tandem repeats characteristic of centromeres, and is, therefore, the putative centromere of *P. falciparum* chromosome 12. To predict the genes encoded by *P. falciparum* chromosome 12, we used 'gene-calling' software in parallel with our colleagues at the WTSI². *Plasmodium falciparum* chromosome 12 is predicted to encode 529 genes (Table 1, and Supplementary Fig. 2), including 23 genes from known *Plasmodium*-specific protein-encoding gene families (eight *vars*, twelve *rifs*, and three *stevors*³) and three transfer RNA genes. The segmental (G + C) content affects the speed and accuracy of sequencing. The predicted exons are, on average, 23.8% (G + C), which is significantly higher than the overall average (19.3%). The predicted introns are, on average, 13.4% (G + C), which is significantly lower than the overall average. All of the chromosome 12 numbers in Table 1 are in accord with the equivalent numbers for the other 13 *P. falciparum* chromosomes^{2,4}.

Independent data support some of the 526 *P. falciparum* chromosome 12 predicted protein-encoding genes/exons, although support for an exon does not necessarily validate the entire gene predicted to contain that exon. Of the 526 predicted genes, 174 (33%) have good matches to sequences in GenBank, while 256 (48.7%) have excellent matches to expressed sequence tags (ESTs, which are short sequences derived from messenger RNAs). In two accompanying publications, Florens *et al.*¹¹ and Lasorder *et al.*¹² report the *P. falciparum* proteome (the sum of all proteins encoded by the *P. falciparum* genome) at the main stages of the complex *P. falciparum* life cycle. Peptides were identified for 268 (51.0%) of the predicted protein-coding sequences of chromosome 12. Our colleagues at the WTSI assigned Gene Ontology (GO) categories to the predicted *P. falciparum* genes² (Table 1).

Table 1 Summary of relevant features

Feature	Value	
	Whole genome	Chr. 12
The genome		
Size (bp)	22,853,764	2,271,477
No. of gaps*	93	0
Coverage†	14.5	14.9
(G + C) content (%)	19.4	19.3
No. of genes‡	5,268	526
Mean gene length (bp)	2,283.3	2,303.1
Gene density (bp per gene)	4,338.2	4,318.4
Per cent coding§	52.6	53.3
Genes with introns (%)	53.9	51.1
Genes with ESTs (%)	49.1	48.7
Gene products detected by proteomics¶ (%)	51.8	51.0
Exons		
Number	12,674	1,270
Mean no. per gene	2.4	2.4
(G + C) content (%)	23.7	23.8
Mean length (bp)	949.1	953.9
Total length (bp)	12,028,350	1,211,430
Introns		
Number	7,406	744
(G + C) content (%)	13.5	13.4
Mean length (bp)	178.7	172.9
Total length (bp)	1,323,509	128,665
Intergenic regions		
(G + C) content (%)	13.6	13.6
Mean length (bp)	1,693.9	1,703.6
RNAs		
No. of tRNA genes	43	3
No. of 5S rRNA genes	3	0
No. of 5.8S, 18S and 28S rRNA units	7	0
The proteome		
Total predicted proteins	5,268	526
Hypothetical proteins††	3,208	332
InterPro matches	2,650	256
Pfam matches	1,746	222
Gene Ontology		
Process	1,301	142
Function	1,244	125
Component	2,412	246
Targeted to apicoplast	551	58
Targeted to mitochondrion	246	32
Structural features		
Transmembrane domain(s)	1,631	155
Signal peptide	544	49
Signal anchor	367	33

EST, expressed sequence tag. Specialized searches used the following programs and databases: InterPro¹⁸; Pfam¹⁹; Gene Ontology²⁰. Predictions of apicoplast and mitochondrial targeting were performed using TargetP²¹ and MitoProtII²²; transmembrane domains, TMHMM²³; and signal peptides and signal anchors, SignalP-2.0²⁴.

*Most gaps are probably <2.5 kb.

†Average number of sequence reads per nucleotide.

‡70% of these genes had similarity to expressed sequence tags or encoded proteins detected by proteomics analyses^{11,12}.

§Excluding introns.

¶Per cent of proteins detected in parasite extracts by two independent proteomic analyses^{11,12}.

††Hypothetical proteins are proteins with insufficient similarity to characterized proteins in other organisms to justify provision of functional assignments.

When sequencing recombinant DNAs/YACs, the possibility always exists that the recombinant DNAs/YACs do not represent accurately the sequence of the original DNA. Bases may have been added, subtracted and/or changed during the biochemical construction of the YACs, and mutations may have occurred during passage of the YACs in yeast. We can address this issue for three of the *P. falciparum* strain 3D7 YACs (341, 293 and 25; Supplementary Table 1), because we had shotgun sequenced these three YACs to high coverage (5.7-, 6.3- and 8.4-fold YAC coverage, respectively; Supplementary Table 1). We separately assembled these three YACs using only the YAC-derived reads, and identified regions of high-quality, well-supported assembled sequence. Then, using the software *cross_match*¹³, we compared the YAC-derived consensus sequence with the overall consensus sequence. From a comparison of a total of 94,151 bp from the three YACs, we found two separate single-base differences. Thus, the resulting frequency of difference between YAC sequence and chromosome sequence is 2 bp/94,151 bp, or 0.000021. Of the three strain B8 YACs that are part of the chromosome 12 tiling path, we sequenced only YAC B8-420 to high YAC coverage (12.9-fold YAC coverage; Supplementary Table 1). We assembled solely YAC B8-420 reads, and identified regions of high quality, well-supported assembled sequence. These regions encompass a total of 43,375 bp. Again, using the software *cross_match*, we compared the high-quality strain B8 sequence to our chromosome 12 consensus sequence over the same 43,375 bp. We found 56 differences of several types, including single-base differences and small deletions/insertions. The resulting DNA polymorphism frequency between the *P. falciparum* strain 3D7 sequence and the strain B8 sequence is 56 bp/43,375 bp, or 0.0013. This frequency is 61 times greater than the mutation frequency (0.0013/0.000021 = 61). □

Methods

DNA sequencing

Plasmodium falciparum chromosome 12 DNA twice purified by contour-clamped homogeneous electric field (CHEF) gel electrophoresis, *P. falciparum* genomic DNA for use as template in PCR reactions, and the appropriate PCR reaction conditions using HotStar kits (Stratagene) were supplied by D. Carucci and his team at the Navy Medical Research Center (NMRC). Yeast/YAC stocks and relevant information were supplied by J. Thompson of the Walter and Eliza Hall Institute (WEHI). The yeast/YACs were grown as described⁷. Agarose plugs containing the YACs were prepared. YACs were twice purified by CHEF gel electrophoresis. The first CHEF gel was composed of standard agarose. The second CHEF gel was composed of low-melting-point (LMP) agarose. YACs were freed from the LMP agarose by agarase digestion at 37 °C. For the construction of shotgun sequencing libraries, the *P. falciparum* chromosome 12 and YAC DNAs were first point-sink sheared (a random shearing process) to an average size of 1 kb for the M13-based vector and 2 kb for the pUC-based vector, as previously described¹⁴. Both M13-based and pUC-based sequencing libraries were constructed from the *P. falciparum* chromosome 12 DNA. Only M13-based sequencing libraries were constructed from the YAC DNAs.

Software

The public software phred was used to call the bases and to assign a quality score to each base^{13,15}; a 'phred score' of 20 or higher is considered good quality. All of the sequence data presented here refer solely to good-quality sequence. The public software phrap was used to assemble the shotgun reads¹⁵. Consed was used to edit the assembled sequence¹⁶. The final gene set was chosen through a manual review of the data. Each base in the open reading frames (ORFs) of the *P. falciparum* chromosome 12 consensus sequence is supported by, at least, three good-quality reads with, at least, one read in each direction. However, because of resource limitations, there are still a few regions (mostly in stretches of repeated sequences) supported by reads in only one direction.

Finishing

There were two types of gaps in the assembled sequence. (1) Plasmid bridges (also known as 'sequence gaps'). A plasmid bridge connects two contigs, and is composed of paired, opposing reads wherein one read is in one contig while the other read is in a second contig. To sequence across plasmid bridges, we designed custom primers in both directions. Using those primers and the particular plasmid as template, we performed primer-extension sequencing. When necessary, we designed custom primers for an additional round of primer-extension sequencing. The addition of primer-extension reads often attracted previously unassembled shotgun reads to that position. (2) Physical gaps. We used two strategies to close physical gaps. The first was to use existing templates that pointed into a physical gap. We designed custom primers and coupled these with their respective templates for primer-extension sequencing. This procedure extended good-quality sequence into a gap. In those cases where there was still more length to the templates

pointing into a gap than covered by good-quality sequence, we again designed custom primers and undertook additional rounds of primer-extension sequencing. When the primer-extension procedure failed to close a physical gap or could not be used because no templates pointed into a gap, we turned to the second, PCR-based, strategy. We designed three nested pairs of custom primers across each physical gap. We used the primer pairs, along with total *P. falciparum* genomic DNA as the template, for PCR reactions. The PCR products were gel-purified and sequenced. Using these two strategies separately or in combination, we were successful in closing every sequence and physical gap in our *P. falciparum* chromosome 12 sequence.

In addition to the gaps, some regions in the assembled sequence of chromosome 12 had good-quality reads in only one direction. Both directions are required, because the sequence in one direction is a check on the sequence in the complementary direction. Therefore, achieving good-quality sequence reads in both directions was a high priority. Where templates existed in the opposing direction, we designed custom primers and undertook primer-extension sequencing on those templates. Where templates did not exist in the opposing direction, we used two different strategies to achieve sequence in the missing direction. One strategy was to undertake an M13 template-based procedure with the existing templates. For this procedure, we started with an M13-based template and used PCR to synthesize the complementary strand in the opposing direction. Then, we sequenced that new DNA strand using primer-extension chemistry. This procedure is often called 'M13-reverses'. The second strategy to achieve sequence in the missing, opposing direction was to construct one or more PCR products across the region. The once-missing, opposing strand of the PCR product was sequenced.

There were many other places in the assembled sequence of chromosome 12 where the sequence was thin (supported by only a few shotgun reads), or ambiguous, or of low quality, and so on. For example, the sequences on both sides of homopolymers of adenine, which occur frequently on this very (adenine + thymine)-rich DNA, were often of low quality. Replacing those thin, weak or ambiguous sequences with good-quality sequence was part of the finishing process. We manually scanned along the entire sequence of chromosome 12, examining both the quality and number of the individual reads and the quality of the consensus sequence. Wherever that quality was low, thin or ambiguous, we designed custom primers for the existing templates. The primers were paired with their appropriate templates for primer-extension sequencing. When this procedure failed, or when there were regions of poor-quality sequence on both strands, we constructed PCR products across the regions and sequenced these PCR products.

PCR products

Because of the very high (A + T) content of *P. falciparum* DNA, the annealing and extension temperatures for PCR reactions are significantly lower (and the extension time significantly longer) than the usual PCR reactions. These lower temperatures might allow slightly mismatched primer/template combinations to be stable and, therefore, amplified. In addition, because of the cost, finishing primers were not purified, so that oligonucleotides of related sequences might be present as contamination in the primer preparations. These related primers might have reasonable matches in the very complex *P. falciparum* genomic DNA template and, therefore, could contribute unwanted primer/template combinations that could be amplified. Therefore, we often found that the products of our PCR reactions were one major DNA product and several minor DNA products, as seen on agarose gels after electrophoresis. As such combinations of DNA do not sequence cleanly, all PCR products to be sequenced were LMP gel-purified.

Annotation

As part of the automated annotation process, the sequences of apparent ORFs were compared to the sequences in GenBank, using the BLAST program¹⁷. Positive quantitative results were posted. Then, we undertook an experiment in community annotation by inviting the world-wide scientific community to enter our website and annotate any particular ORF, or gene, or gene family, of their choice. At the time of writing, 18 scientists have annotated 52 genes. The participating annotators are: A. Danchin, C. Doerig, A. H. Fairlamb, P. Horrocks, J. E. Hyde, G. Plunkett, S. Rahlfs, P. Rathod, P. A. Rea, M. Seaman, C. Slomianny, J. Tyler, J. Kadonaga, C. Vaquero, C. Boschet, J. Vinetz, L. Wilming and M. F. Wiser. This pilot experiment in community annotation has been a modest, but real, success.

Received 14 June; accepted 9 September 2002; doi:10.1038/nature01102.

1. Berman, J. G. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**, 1–11 (2001).
2. Hall, N. et al. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature* **419**, 527–531 (2002).
3. Gardner, M. J. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
4. Gardner, M. J. et al. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* **419**, 531–534 (2002).
5. Gardner, M. J. et al. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
6. Bowman, S. et al. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
7. Rubio, J. P., Thompson, J. K. & Cowman, A. F. The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO J.* **15**, 4069–4077 (1996).
8. Su, X. et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
9. Su, X. Z. & Wellems, T. E. *Plasmodium falciparum*: assignment of microsatellite markers to chromosomes by PFG-PCR. *Exp. Parasitol.* **91**, 367–369 (1999).
10. Jing, J. et al. Optical mapping of *Plasmodium falciparum* chromosome 2. *Genome Res.* **9**, 175–181 (1999).
11. Florens, L. et al. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).

12. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).
13. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
14. Oefner, P. J. *et al.* Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res.* **24**, 3879–3886 (1996).
15. Ewing, B., Hillier, L., Wendt, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
16. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
18. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
19. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
20. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
21. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
22. Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
23. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
24. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We acknowledge the generosity of the participating scientists at Stanford University, TIGR, the WTSI, the NMRC and Oxford University. We also thank N. Hall, M. Berriman, A. Pain and B. Barrell for their time and expertise during the gene-calling annotation process, and are grateful to the members of our Stanford Genome Technology Center for their assistance throughout this project. We thank the Burroughs Wellcome Fund for support that allowed us to participate in the international Malaria Genome Project.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W.H. (e-mail: hyman@sequence.stanford.edu). The GenBank accession number of the sequence of *P. falciparum* (clone 3D7) chromosome 12 is AEO14188.

Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry

Edwin Lasonder^{*†}, Yasushi Ishihama^{*}, Jens S. Andersen^{*},
Adriaan M. W. Vermunt[†], Arnab Pain[‡], Robert W. Sauerwein[§],
Wijnand M. C. Eling[§], Neil Hall[‡], Andrew P. Waters^{||},
Hendrik G. Stunnenberg[†] & Matthias Mann^{*}

^{*} Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

[†] Department of Molecular Biology, NCMLS, University of Nijmegen, Geert Grooteplein 26, 6525 GA Nijmegen, The Netherlands

[‡] The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

[§] Department of Medical Microbiology, NCMLS, University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands

^{||} Leiden Malaria Research Group, Department of Parasitology, Centre for Infectious Disease, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

The annotated genomes of organisms define a 'blueprint' of their possible gene products. Post-genome analyses attempt to confirm and modify the annotation and impose a sense of the spatial, temporal and developmental usage of genetic information by the

organism. Here we describe a large-scale, high-accuracy (average deviation less than 0.02 Da at 1,000 Da) mass spectrometric proteome analysis^{1–3} of selected stages of the human malaria parasite *Plasmodium falciparum*. The analysis revealed 1,289 proteins of which 714 proteins were identified in asexual blood stages, 931 in gametocytes and 645 in gametes. The last two groups provide insights into the biology of the sexual stages of the parasite, and include conserved, stage-specific, secreted and membrane-associated proteins. A subset of these proteins contain domains that indicate a role in cell–cell interactions, and therefore can be evaluated as potential components of a malaria vaccine formulation. We also report a set of peptides with significant matches in the parasite genome but not in the protein set predicted by computational methods.

The *Plasmodium falciparum* parasite is pre-committed to one of three different developmental pathways on re-invasion of a host erythrocyte⁴. Either it develops asexually, resulting in proliferation, or it develops into a male or a female gametocyte—sexual precursor forms that maintain a stable G1 cell cycle arrest and circulate in the peripheral blood stream when mature. Gametocytes are activated in the mid-gut of the mosquito when ingested by the vector through the consumption of a blood meal, and rapidly develop into mature gametes that fertilize to form zygotes. The zygotes mature into a motile invasive form, the ookinete, which is adapted for colonization of the mosquito. Clearly, sexual development and fertilization are essential processes within the parasite life cycle and one strategy of vaccination (transmission blocking) seeks their interruption through antibody-based blockades. Although good candidates for such vaccines exist, it is an accepted view that an effective vaccine will need to target several stages of the parasite and several components of the different forms of the parasite⁵. High-throughput proteome studies on pure parasite forms are a rapid and sensitive means to discover such vaccine candidates.

To define the proteome of the asexual and sexual blood stages of the malaria parasite *P. falciparum* (NF54 isolate), purified asexual (trophozoites and schizonts, Fig. 1a, left panel) and sexual stage parasites (gametocytes, right panel) or gametes (not shown) were extracted by freeze–thawing and centrifugation, yielding soluble and insoluble (pellet) fractions (Fig. 1b). The result of a typical gametocyte extraction is shown in Fig. 1c, revealing that most of the *P. falciparum* and red blood cell (RBC) proteins were present in the soluble fraction, whereas membrane proteins such as the gametocyte-specific cell surface protein Pfs48/45 (ref. 6) were found exclusively in the pellet fraction, as revealed by western blotting (Fig. 1c). These complex protein mixtures were then analysed 'gel free' (differentially extracted membrane fractions) or separated into ten molecular mass fractions by one-dimensional gel electrophoresis followed by excision of equally spaced bands after precisely removing haemoglobin and globin (Fig. 1c, right panel), and tryptic digestion. The tryptic peptides were separated by reversed phase liquid chromatography coupled to quadrupole time-of-flight mass spectrometry for peptide sequencing (nanoLC-MS/MS). Iterative calibration algorithms were used to achieve a final, average absolute mass accuracy of better than 20 parts per million (p.p.m.) in both the precursor and fragment ions, or a mass deviation of 0.03 Da for a typical tryptic peptide of mass 1,300 Da.

These high-accuracy spectra were searched against a combined human and draft *P. falciparum* database, using probability-based scoring in which the fragment ions are matched against the calculated fragments of all tryptic peptides from the human and parasite sequences⁷. A total of 7,548 distinct peptides from the putative set of malaria proteins were matched with significant probability scores (Supplementary Table A). These peptides mapped to 1,709 malaria proteins. Additional constraints were applied to the peptides, including peptide size, discrimination to the next best match, and features of the tandem mass spectra such as

the dominant amino-terminal and inhibited carboxy-terminal fragmentation of proline. These criteria resulted in a list of 1,289 unique proteins that were identified with high confidence, comprising about 23% of the current proteome prediction⁸ (Supplementary Table B). Roughly one-third of the malaria proteins identified appeared in both the soluble and insoluble fractions; the remaining two-thirds were present either in the insoluble or in the soluble fraction (Fig. 2a). Taken together, the analysis revealed 714 malaria proteins in asexual stage parasite preparations, 931 in gametocytes and 645 in gametes. The analyses also resulted in identification of more than 1,000 human proteins.

At present, 315 malaria proteins have been found solely in gametocytes, 226 in trophozoites and schizonts, and 97 in gametes (Fig. 2b). A total of 575 proteins were found only in sexual stages (gametocytes plus gametes) and 488 proteins were present in both asexual and sexual stages. The definitive classification of a protein as stage-specific is tentative however, and awaits full analysis of all life cycle stages. Moreover, biological samples are rarely completely pure or synchronous (see Methods). The well-characterized, sexual stage-specific surface antigen, Pfs48/45, highlights this point. Three Pfs48/45-derived peptides were detected in the asexual preparations as opposed to 57 in gametocytes. Integration of the peptide ion currents, a quantity roughly proportional to molar

amount of peptide species (see Supplementary Fig. A), yielded a protein abundance ratio of more than eight between the two stages. The unexpected presence of Pfs48/45 in asexual stages is probably due to a small percentage of committed gametocytes in the asexual preparations from previous cycles of growth that cannot be physically separated from asexual-stage parasites. This interpretation is supported by western and northern blot analyses⁹ and was further corroborated by quantitative polymerase chain reaction with reverse transcription (RT-PCR) analysis (Fig. 1c and Table 1). Notwithstanding the restrictions, the data are of sufficiently high quality and coverage to justify the classification of a large number of known and new proteins as 'stage-specific'. In fact, the small gametocyte contaminations in the asexual preparation strengthens the stage-specific classification of those proteins that are found exclusively in sexual stage parasites such as Pfg377, gene 11-1 and the hypothetical protein PF14_0039. A total of 517 peptides from the sexual stage antigen Pfg377 (ref. 10) were detected in the gametocytes and gamete preparations whereas not a single Pfg377 peptide was detected in the asexual stage, illustrating the exquisite sensitivity of the analytical procedure as well as the quality of the biological sample. Collectively, mass spectrometric and quantitative RT-PCR analysis (see below) indicates that expression of Pfg377, gene 11-1 and the hypothetical protein PF14_0039 is turned on later

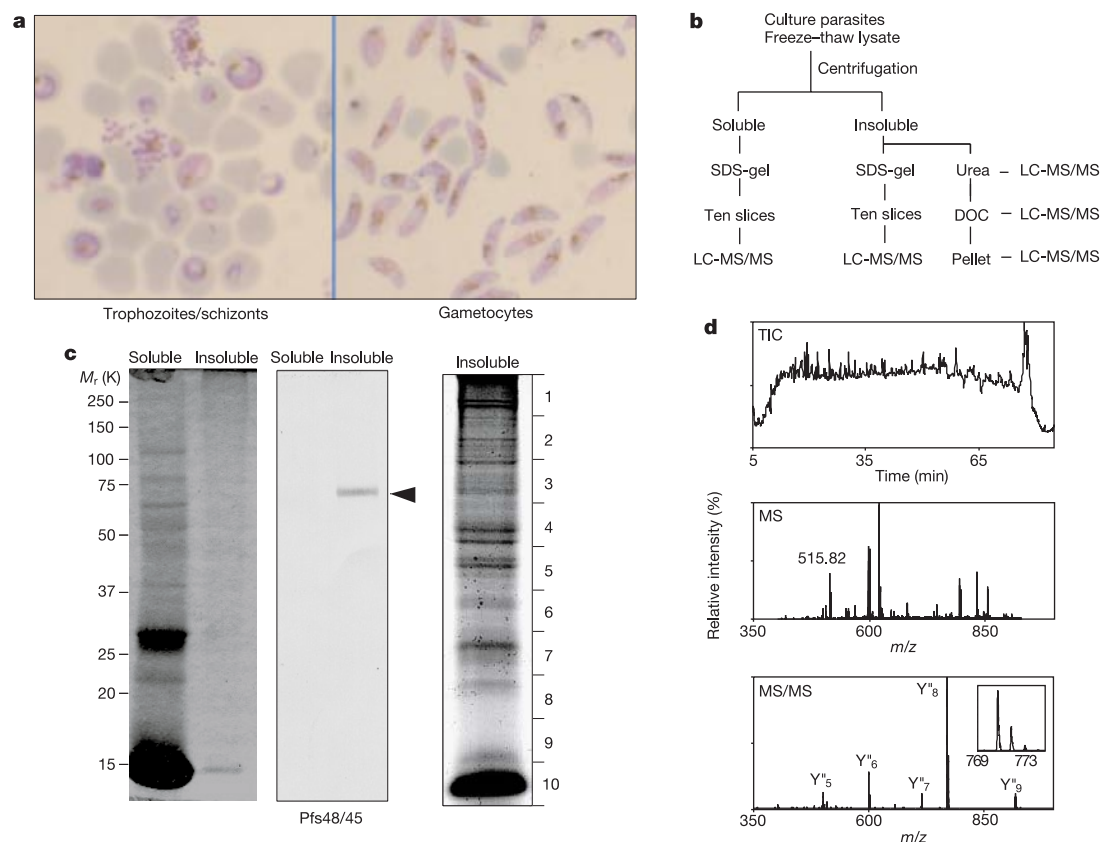


Figure 1 Differential extraction of parasite-infected red blood cells (RBCs) and flow chart of MS analysis. **a**, Giemsa staining of purified trophozoites and schizonts (left panel; asexual stage parasites), and gametocytes (right panel; sexual stage parasites). **b**, Extraction procedure for infected RBCs using freeze-thaw lyses and centrifugation, yielding a soluble and insoluble (pellet) fraction. **c**, The soluble and insoluble fraction from 5×10^5 gametocytes were separated on a 10% SDS gel and stained with Coomassie blue (left panel) or processed for western blotting and developed with a rabbit polyclonal antibody to Pfs48/45. The arrowhead indicates Pfs48/45. The right hand panel shows the

insoluble fraction of 2×10^7 gametocytes. **d**, Mass spectrometric analysis of one of the gel slices shown in **c**. The top panel shows the summed ion current of all peptides eluting at a particular time. The middle panel shows a typical mass spectrum obtained from the eluting peptides. The bottom panel shows a fragmentation spectrum of the peptide with a mass of 515.82 indicated in the middle panel and magnification (inset) of one of the fragments. Database searching with this fragmentation information results in identification of peptide LFGIVGSIK from Pfs48/45.

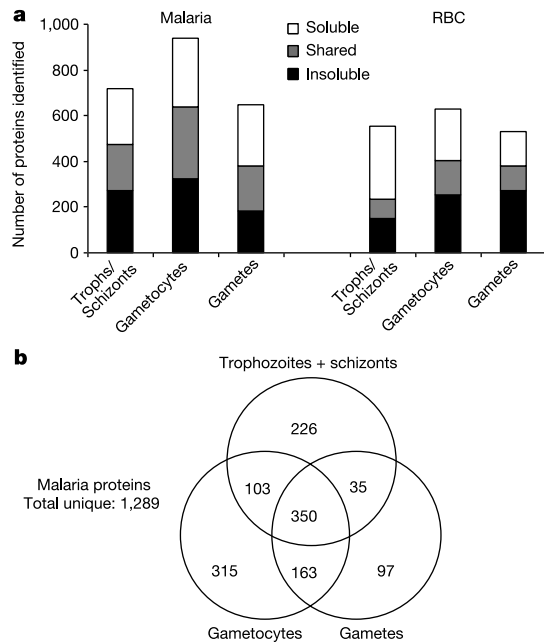


Figure 2 Schematic representation of proteomic data. **a**, Compilation of unique and common proteins in the soluble and insoluble fractions. Bars represent the number of proteins in the different preparations (trophozoites/schizonts, gametocytes and gametes) that were assigned to malaria or human RBCs. Proteins present solely in soluble (open bars) or insoluble (black bars) fractions as well as protein common to both fractions (shaded bars) are indicated. **b**, Venn diagram of the distribution of identified malaria proteins over the three blood stages. The distribution between trophozoites plus schizonts, gametocytes and gametes is shown.

during gametogenesis than Pfs48/45.

Among the identified asexual proteins described previously are stage-specific antigens KAHRP, PfEMP3, PfSAR1 and PfSBP1, which are involved in the assembly of structures specific to the blood stage, such as knobs¹¹. Schizont-specific proteins associated with the merozoite surface and invasive organelles such as MSP1, -2, -3, -6, -7 and -7a, or AMA1 and rhoptry-associated proteins RAP1, RAP2 and Rhop1 (for review of the proteins involved, see ref. 12)

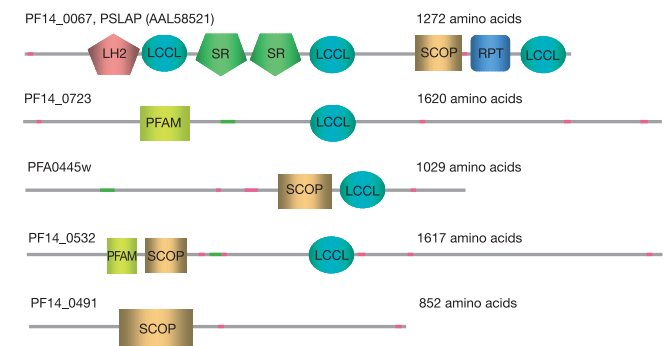


Figure 3 LCCL/lectin domain proteins expressed in sexual stages of *P. falciparum*. The SMART²²-generated graphic descriptions of the five proteins expressed in sexual stages that contain LCCL/lectin domains are shown. The *P. falciparum* homologue of PSLAP¹⁶ (GenBank accession number AAL58521), PF14_0067, contains a signal sequence (cleavage: amino acids, aa, 22 and 23); 3 LCCL domains (position: aa, 275–362, E value 1.28×10^{-3} , 670–676, E value 8.11×10^{-5} , and 1169–1255, E value 6.89×10^{-16}); 2 scavenger receptor (SR) domains (aa 401–515, E value 3.41×10^{-6} and 528–642, E value 3.10×10^{-7}); 1 lipoxigenase homology 2 (LH2) domain (aa 157–265, E value 8.20×10^{-3}); and one SCOP ConA lectin-like domain (aa 911–1013, E value 1×10^{-4}). PF14_0723 contains a signal peptide (cleavage: aa 19 and 20); 1 LCCL domain (aa 752–843, E value 1.4×10^{-21}); and a PFAM predicted discoidin domain (aa 296–420, E value 8.7×10^{-2}). PFA0445w contains a predicted signal sequence (cleavage: aa 24 and 25); 1 LCCL domain (aa 740–827, E value 4.49×10^{-6}); 1 SCOP kringle family domain (aa 43–95, E value 1.61×10^{-2} , not shown); and 1 SCOP dld7pm galactose-binding motif (aa 593–714, E value 4.01×10^{-3}). PF14_0532, also recognized as a gametocyte-specific transcript in *P. berghei* (AF491294), contains signal sequence (cleavage: aa 23 and 24); 1 LCCL domain (aa 724–815, E value 8.87×10^{-4}); and 1 PFAM ricin domain (aa 222–269, E value 4.0×10^{-2}). PF14_0491 is not predicted to contain a signal sequence but contains a SCOP dlacc anthrax protective antigen family with an immunoglobulin fold (aa 204–372, E value 3×10^{-9}) and SCOP/FN2 kringle-like (aa 30–80, E value 6.7×10^{-2} , not shown).

were also readily detected. Additionally, numerous hypothetical proteins containing a predicted signal sequence (61) and/or a transmembrane domain (189) were discovered, thus extending the repertoire of confirmed gene products and of potential new vaccine candidates (Supplementary Table B). Our analysis did not reveal any high-scoring peptides in the asexual blood stage preparations that could be assigned unambiguously to a member of the highly variable surface molecules of the PfEMP1 family, which are

Table 1 Quantification of stage-specific proteins

Protein name	Accession no.	RT-PCR*	MS-XIC†
Asexual stage-specific proteins (known)			
Apical membrane antigen 1 (AMA1)	PF11_0344	3 (1)	>2
Cyto-adherence-linked asexual protein (CLAG9)	PF11730w	13 (3)	>5
Mature parasite-infected erythrocyte surface antigen (MESA)	PFE0040c	7 (4)	>12‡
Merozoite surface protein 1 (MSP1)	PF11475w	66 (37)	>60
Cysteine protease	PFB0340c	221 (61)	>8
Sexual stage-specific proteins (known)			
Actin II	PF14_0124	26 (4)	>10
Pfs16 surface antigen	PF11_0318	8 (2)	>6
Sexual stage-specific surface antigen (Pfs48/45)	PF13_0247	43 (19)	8.2 (4.0)‡
Transmission-blocking target antigen (Pfs230)	PFB0405w	61 (20)	>6
<i>Plasmodium falciparum</i> gametocyte antigen (Pfg377)	PFL2405c	289 (79)	>20
Gene 11-1	PF10_0374	160 (14)	>8
Sexual stage antigen Pfs25	PF10_0303	858 (290)	>4
Sexual stage-specific proteins (novel)			
Hypothetical protein	PF14_0039	597 (208)	>17
Hypothetical protein	PF11_0310	9 (5)	>2
Hypothetical protein	PF11_0413	5 (1)	2.4 (0.5)

*Ratio between stages. RT-PCR normalized to heat shock protein 86, HSP86. Standard deviations are indicated in parentheses.

†Ratio between stages. Values were calculated from ion currents for the precise peptide molecular mass using maximum peak heights of peptides with more than 30 ion counts.

‡Extracted ion chromatogram (XIC) of the peptides of these proteins are presented in Supplementary Fig. A.

knob-associated and primarily responsible for cyto-adherence of parasites to the peripheral vascular endothelium. PfEMP1 may be expressed at very low abundance in *P. falciparum* strain NF54, or poorly extracted by our methods. Our mass spectrometry analyses revealed almost all of the known proteins specific to *P. falciparum* gametocytes and gametes (Supplementary Table B). For example, four of the six members of the P48/45 family that are transcribed in sexual stages were readily identified. In addition we detected the sexual stage-specific Pfs48/45 paralogue, Pf47, for the first time.

To corroborate and extend the stage specificity, we compared not only the absence or presence of particular peptides in the respective fractions, but also integrated the ion currents of peptides (see Supplementary Fig. A). This is important because it is possible that a peptide is not selected for sequencing owing to the complexity of the sample, but is nevertheless present at significant amounts. For example, MESA was found with 18 peptides in the asexual stage. The ion currents at the corresponding elution times in the preparation of a sexual stage were inspected but did not show any signal at the respective peptide masses. Similarly, analysis of peptide ion currents

corresponding to hypothetical proteins such as PF14_0039 revealed a protein abundance ratio of at least 17:1 between the sexual and asexual stages. We next performed quantitative RT-PCR using messenger RNA from parallel asexual, gametocyte and gamete parasite preparations and gene-specific primer sets for a group of hypothetical and known proteins and an arbitrary selection of putative sexual and asexual stage-specific proteins. RT-PCR ratios of signals obtained in asexual versus sexual parasites ranged from 3-fold to over 850-fold (Table 1). In accord with the mass spectrometric findings, mRNA from Pfg377, gene 11-1 and the protein PF14_0039, but not Pfs48/45, were detected exclusively in gametocyte and gamete preparations. Comparison between the results obtained with mass spectrometry and quantitative RT-PCR shows that the changes observed in protein abundance between the asexual and sexual stages were also reflected in their mRNA levels.

Erythrocytic stages of the parasites are under enhanced oxidative stress and are particularly vulnerable to exogenous challenges by reactive oxygen species. Therefore it is interesting to note that proteins involved in protection against oxidative stress—the glyoxalase enzymatic system associated with glutathione—seem to be upregulated in gametes and gametocytes, such as glyoxalases I (PF11_0145) and II (PFL0285w), and two different glutaredoxins (MAL6P1.72 and PFC0271c)¹³. Some of the enzymes involved in the synthesis and metabolism of glutathione such as glutathione S-transferase (PF14_0187) and γ -glutamylcysteine synthetase (PFI0925w) are exclusively present in sexual stages. Others were found in sexual as well as asexual stages; these include glutathione peroxidase (PFL0595c) and glutathione reductase (PF14_0192). Furthermore, thioredoxin systems are also readily found in sexual stages (thioredoxin, PF14_0545 and PF13_0272 (secreted); thioredoxin peroxidase, PFL0725w; and thioredoxin reductase, PFL0725w).

Another stage-specific feature is the motility of male gametes. This involves the axoneme, a specialized structure consisting of two parallel microtubules thought to be connected by dynein complexes, which so far have only been demonstrated in mature merozoites¹⁴. The full *P. falciparum* genome reveals 12 genes that are expected to encode different dynein forms. They are conserved throughout *Plasmodium* species and have maximum homology to flagellar dyneins (for example, PF11_0240). Our analysis indicates that at least six of these (three light and three heavy chains) are expressed exclusively in sexual stages. Further specialized motility-associated proteins that may assist in the formation of the axoneme, such as actin II, α -tubulin II, β -tubulin, putative kinesins, as well as predicted tubulin-specific chaperones, are also evident.

The male gamete has been shown to bind to sialic acid on erythrocytes through a lectin-like activity¹⁵. Therefore, one anticipated finding within the proteome of the sexual stage was the presence of proteins (protein families) that exhibit appropriate adhesion properties. Our analysis revealed five proteins expressing predicted lectin domains of which four contain an additional LCCL motif thought to be involved in protein–protein interactions—all five are expressed exclusively in sexual stages (Fig. 3). One of these proteins, PSLAP (PF14_0067; AAL58521), has been characterized previously as gametocyte-specific and has been shown to contain multiple different domains suggestive of a role in cell–cell interactions¹⁶. A second (PF14_0723) has been characterized as a gametocyte-specific transcript in the rodent parasite *Plasmodium berghei* (GenBank accession number AF491294). Four of the five are predicted to be secreted proteins (PF14_0723, PFA0445w, PF14_0532, PSLAP). A fifth protein (PF14_0491) contains only a lectin domain and is predicted to be non-secreted; however, we believe that this is due to annotation based on a short exon-splice model, because alternative models would create a signal peptide. Interrogation of the *Plasmodium* genomes (*P. falciparum* and *P. y. yoelii*) suggests that these proteins are conserved in both models

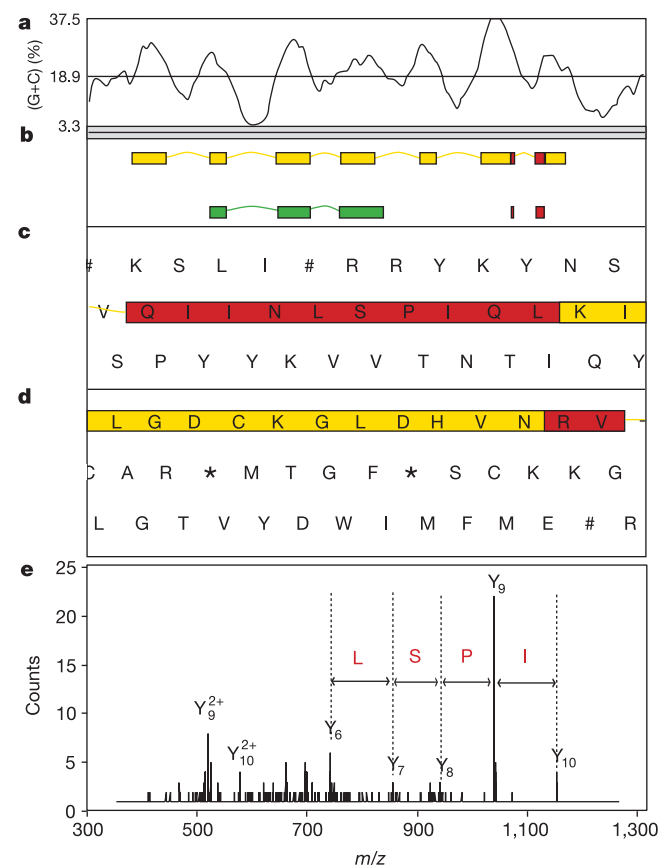


Figure 4 Refinement of gene structure and re-annotation using proteomic data. **a**, (G + C) plot in an ARTEMIS window corresponding to the genomic region of the gene PF11_0169 on chromosome 11 of *P. falciparum*. **b**, Alternative gene models for PF11_0169. The green and yellow gene models represent the original and refined gene structure of PF11_0169, respectively. The red blocks represent the peptide LQIPSLNIIQVR and its occurrence in respect to the green and the yellow gene models. **c**, **d**, Sections from an ARTEMIS window showing the splice-donor site of exon I and the splice-acceptor site of exon II of the yellow gene model. The three-frame translation is shown with hash or asterisk, denoting stop codons. The red blocks together represent the full-length peptide LQIPSLNIIQVR. **e**, Tandem mass spectrum of the orphan peptide LQIPSLNIIQVR. C-terminal fragment ions (Y ions) are indicated as well as a partial sequence.

and in other *Plasmodium* species that infect humans, and that expression of LCCL-domain-containing proteins is restricted to the parasite stages found in the mid-gut of the mosquito. Future investigations will determine the precise role of these proteins in the complexity of gamete fertilization, erythrocyte binding by male gametes, interactions with the host and vector immune systems, and the surface carbohydrate structures of the mosquito mid-gut.

The gel slice approach chosen for proteome analysis allowed us to correlate apparent molecular masses with gene annotation derived from theoretical mass predictions, revealing outliers that may be due to protein processing or to incomplete gene annotation. Protein Pfg377, for example, probably represents a case of protein processing (Supplementary Fig. B). The C-terminal portion of Pfg377 was covered with 69 sequenced peptides solely in the protein mixture from gel slice 5 (relative molecular mass (M_r) range of 100,000 to 140,000 (100K–140K)), whereas 15 peptides matching the N-terminal portion were found exclusively in gel 7 (65K–84K). Northern blot analysis¹⁰ and quantitative RT–PCR using 5' and 3' specific primer sets (data not shown) indicate that the proteins are encoded by a single gene that is processed after translation.

The parasite genome is highly (A + T)-rich and has proven difficult to assemble and annotate⁸. For this reason, we also performed direct genome searches¹⁷ where the malaria genome rather than the predicted set of proteins was searched by the mass spectrometric data. A large number of additional hits emerged. After analysing these 'orphan' peptide hits in the genome of *P. falciparum*, we were able to identify additional exons or assign different exon–intron boundaries of previously annotated genes using the ARTEMIS annotation tool¹⁸. One such orphan peptide, LQIPSLNIQVR, identified two additional exons of the gene PF11_0169 on chromosome 11, annotated as a hypothetical protein. With this information we were able to identify two further exons and modify the gene structure appropriately. This led to re-annotation of the previously annotated hypothetical protein PF11_0169 as a putative homologue of sno-type pyridoxin biosynthesis protein (Fig. 4). The orphan peptide not only identified the two exons of the gene PF11_0169 but also verified the boundaries of the intron between the exons I and II. With the refined gene model we were able to re-annotate the gene with appropriate Gene Ontology (GO) terms. Initial analysis similar to the one outlined above led to eight definite and six probable gene annotation changes. Supplementary Table C lists more than 100 high-quality orphan peptides that can be used in gene annotation and the finding of new genes.

We have applied state-of-the-art proteomics techniques to obtain valuable information about stage-specific expression and localization of malaria proteins as well as information useful in confirming and extending the bioinformatics analysis of the proteome. Further work could encompass additional stages of the parasite life cycle and make use of the rapidly evolving mass spectrometric technology, associated bioinformatics, as well as direct comparison of stages using stable isotope techniques. □

Methods

Preparation of parasites

Plasmodium falciparum parasites were cultured using a semi-automated culture system. Asexual stages (trophozoites and schizonts) were purified using the Variomacs system essentially as described¹⁴. For the production of gametocytes, parasites were cultured for 14 days without addition of red blood cells. The gametocyte cultures were treated with 50 mM N-acetyl-glucosamine from day 8 to 12 after the start of the culture to remove asexual stage parasites. After a total of 14 days culture the gametocytes were collected and purified using the Variomacs system as described¹⁴. For the production of gametes, purified gametocytes were pelleted and incubated in G_m buffer containing 100 μ M xanthurenic acid for a period of 3 h at 21 °C (ref. 19). (G_m buffer contains 1.67 mg ml⁻¹ glucose, 8 mg ml⁻¹ NaCl, 1 mg ml⁻¹ Tris, pH 8.1.) Every effort was made to minimize enzymatic activity and protein degradation during sampling and the subsequent isolation of the parasites. However, we cannot exclude that some of the differences in protein

profiles that we observe between the different life-cycle stages may be a consequence of the sample-handling procedures.

Sample preparation

The infected red blood cells (10^7 cells) of each stage were divided into a soluble and insoluble fraction by freeze–thawing several times and pelleted by centrifugation at 13,000 r.p.m. (16,100 g). Proteins were extracted in SDS–polyacrylamide gel electrophoresis (PAGE) loading buffer and separated into ten fractions on a 10% protein gel. The separated proteins were treated with dithiothreitol (DTT) and iodoacetamide, and in-gel digested with trypsin²⁰. Proteins were also extracted by 8 M urea, 100 mM Tris–HCl, pH 8.0, treated with DTT and iodoacetamide, and digested in-solution with endoprotease Lys-C and trypsin after dilution to 2 M urea. Proteins from the insoluble fractions were further extracted with 1% deoxycholic acid (DOC), 100 mM Tris–HCl, pH 8.0, and digested together with the remaining pellet as above in the presence of 0.05% DOC.

NanoLC-MS/MS analysis of malaria proteins

Peptide mixtures were loaded onto 75- μ m ID columns packed with 3- μ m C18 particles (Vydac) and eluted into a quadrupole time-of-flight mass spectrometer (QSTAR, Sciex–Applied Biosystems). Fragment ion spectra were recorded using information-dependent acquisition and duty-cycle enhancement (see ref. 21 and references therein for a more detailed description of the method). The three malaria stages studied were analysed at least in duplicate. In total, more than 100 nanoLC-MS/MS runs were analysed, yielding more than 200,000 peptide-sequencing events.

Data analysis

Peak lists containing the precursor masses and the corresponding MS/MS fragment masses were generated from the original data file and searched in the annotated *P. falciparum* database (Sanger/TIGR) combined with the human IPI database (European Bioinformatics Institute) using the Mascot program (Matrix Science). Identified peptides that were not unique or had a score less than 20 were removed. Proteins identified with a combined peptide score of higher than 60 were considered significant, and lower scoring proteins were manually verified or rejected. Iterative calibration algorithms on the basis of identified peptides were used to achieve a final average absolute mass accuracy of better than 20 p.p.m. in both the precursor and fragment ions. Relative protein abundance between the malaria stages was based on the total number of unique peptides identified for each protein and was further supported by extracted ion chromatograms (XIC) for individual peptides within an elution time window of 3 min (see also Supplementary Fig. A).

Quantitative real-time RT–PCR

Relative mRNA abundance was quantified by real-time PCR with the GeneAmp 5700 Sequence Detection System (Applied Biosystems) using the SYBR Green PCR Master Mix kit (Applied Biosystems). Total RNA was isolated from two stages: asexual blood stage (trophozoites and schizonts) and the sexual blood stage (gametocytes), both prepared as described above. RNA was isolated from 4×10^7 parasites using TRIzol reagent (Gibco BRL). A total of 2 μ g of RNA was used for complementary DNA synthesis. RNA was treated with DNaseI (Pharmacia) after which cDNA was synthesized with random hexamers and Superscript II enzyme (Gibco BRL), according to standard protocols. cDNA was dissolved in 50 μ l diethyl pyrocarbonate (DEPC)-treated H₂O. A total of 3 μ l of $\times 10$ and $\times 100$ diluted cDNA was used for real-time PCR. Primers were designed by Primer Express software (Applied Biosystems). The mRNA levels for each gene were normalized against heat shock protein 86. For each gene the relative mRNA abundance was determined by calculating the ratio of asexual:gametocyte stage and gametocyte:asexual stage.

Received 31 July; accepted 9 September 2002; doi:10.1038/nature01111.

- Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837–846 (2000).
- Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**, 676–682 (1999).
- Griffin, T. J. & Aebersold, R. Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* **276**, 45497–45500 (2001).
- Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100**, 191–200 (1990).
- Richie, T. L. & Saul, A. Progress and challenges for malaria vaccines. *Nature* **415**, 694–701 (2002).
- Kocken, C. H. *et al.* Cloning and expression of the gene coding for the transmission blocking target antigen Pf48/45 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **61**, 59–68 (1993).
- Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
- Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
- van Dijk, M. R. *et al.* A central role for Pf48/45 in malaria parasite male gamete fertility. *Cell* **104**, 153–164 (2001).
- Alano, P. *et al.* COS cell expression cloning of Pfg377, a *Plasmodium falciparum* gametocyte antigen associated with osmiophilic bodies. *Mol. Biochem. Parasitol.* **74**, 143–156 (1995).
- Wickham, M. E. *et al.* Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J.* **20**, 5636–5649 (2001).

12. Cowman, A. F. *et al.* Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. *FEBS Lett.* **476**, 84–88 (2000).
13. Rahlfs, S., Fischer, M. & Becker, K. *Plasmodium falciparum* possesses a classical glutaredoxin and a second, glutaredoxin-like protein with a PICOT homology domain. *J. Biol. Chem.* **276**, 37133–37140 (2001).
14. Fowler, R. E. *et al.* Microtubule associated motor proteins of *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* **117**, 187–200 (2001).
15. Templeton, T. J., Keister, D. B., Muratova, O., Procter, J. L. & Kaslow, D. C. Adherence of erythrocytes during exflagellation of *Plasmodium falciparum* microgametes is dependent on erythrocyte surface sialic acid and glycophorins. *J. Exp. Med.* **187**, 1599–1609 (1998).
16. Delrieu, I. *et al.* PSLAP, a protein with multiple adhesive motifs, is expressed in *Plasmodium falciparum* gametocytes. *Mol. Biochem. Parasitol.* **121**, 11–20 (2002).
17. Kuster, B., Mortensen, P., Andersen, J. S. & Mann, M. Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics* **1**, 641–650 (2001).
18. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
19. Brooks, S. R. & Williamson, K. C. Proteolysis of *Plasmodium falciparum* surface antigen, Pf5230, during gametogenesis. *Mol. Biochem. Parasitol.* **106**, 77–82 (2000).
20. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
21. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).

22. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank our colleagues for help and discussions. This work was supported by the Danish National Research Foundation, the Dutch Science Foundation (NWO), the European Union and the World Health Organization (WHO) Special Program for Research and Training in Tropical Diseases.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.G.S. (e-mail: h.stunnenberg@ncmls.kun.nl) or M.M. (e-mail: mann@bmb.sdu.dk). Sequence data for the genes newly annotated according to the present study can be found at http://www.sanger.ac.uk/Projects/P_falciparum.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Keyvin Johnson

naturejobs

Career prospects in Europe

“Europe’s forgotten young researchers.” “Pay, pension and benefits: why should scientists be disadvantaged compared with other professions?” With compelling talk titles such as these, a conference in Heidelberg this September, sponsored by the European Molecular Biology Organization (EMBO) and the European Life Science Forum (ELSF), served as an opportunity to pick apart the career structure of European life scientists.

And what conclusions did the 40 or so attendees, who ranged from young investigators to administrators, come to? “The fact that there is no real career structure was the main focus,” says Luc Van Dyck, executive coordinator of ELSF and one of the event’s organizers. Another was “the fact that there is no real tenure-track program similar to the US,” he says.

One of the main problems identified was the variation within Europe, with some countries such as the United Kingdom having more flexible pathways through academia compared to others’ more rigid systems. This disparity discourages mobility within Europe. Combined with lower pay for academics than in the United States or in industry, this creates a “leaky pipeline” where promising European scientists leave for the US or opt for another career.

So what can be done, now that the issues have been identified? EMBO and ELSF will issue recommendations by the end of the year, which will include educating students about careers in science beyond academia. But if universities are to prevent the best European students from slipping away, they will need to pay more and create incentives to enter an academic career, rather than building obstacles. If these changes aren’t made, the next conference will have sessions entitled, even less optimistically perhaps: “Where have all the young scientists gone?”

Paul Smaglik



Naturejobs editor

Contents

SPECIAL REPORT

Genome boost for
malaria research

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

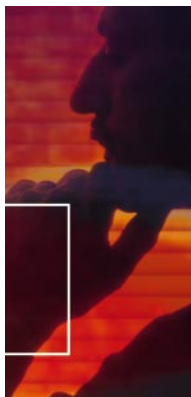
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Will the malaria genome open the door to new drugs, vaccines and job opportunities?

Di Gershon investigates

After Tanzanian-born Pradipsinh Rathod completed his PhD 15 years ago, he wanted to work on a disease relevant to developing countries. He picked malaria because it kills over a million people a year, primarily children under five in sub-Saharan Africa.

In the beginning, he says, it was hard to get a good academic job doing malaria research. "Even parasitologists looked down on it as a difficult system, and not the one you would pick if you were trying to make a career in academia," says Rathod, now professor of the department of chemistry at the University of Washington, Seattle.

The professional outlook for malaria researchers began to change six years ago when four groups got together to sequence the human malaria parasite, which was divided up chromosome by chromosome among them. And it should change even more now that researchers have at their disposal the genomes of all three organisms involved in the transmission of malaria — the parasite *Plasmodium falciparum* (see page 498, this issue) and its mosquito vector *Anopheles gambiae* (*Science* 298, 291; 2002; see box opposite), and the human host, published last year (see *Nature* 409, 860–921; 2002).

This information should help identify the genes and pathways involved in the transmission, pathogenicity and drug responses of malaria, and provide a greater understanding of the complex host–parasite and host–vector interactions. In the longer term, it may reveal new targets for drug development, better targets for vaccines and new ways to interrupt malaria transmission. Funding — although still modest compared to some other pathogens that affect both the developing and developed world — is up, and prospects for a career in malaria research have never looked better.

RETOOLING

The metamorphosis of Rathod's lab, following the release of the raw sequence data, is illustrative. He is completely retooling his lab set-up.

Five years ago, his lab consisted of a bunch of individual scientists, each working on a different gene or a different drug. "We would work endlessly on trying to express a particular protein in abundant amounts and trying to look at the kinetics of binding

of a particular drug to a protein," he says. Now, using microarray technology, for example, it is possible to look at gene expression in the whole organism rather than just individual genes.

As a result, students are now trained more broadly. In Rathod's lab, for example, they would learn some cell biology, molecular biology, organic synthesis, bioinformatics and computer science. This, he says, better prepares them for what's going on — not just in malaria but in many other multidisciplinary fields as well. Moreover, he says, "the field is incredibly welcoming of the types of scientists that we didn't interact with a lot before, such as engineers, computer scientists and organic chemists".

In his department, he says, even people doing work on nanotechnology and microfluidics are becoming interested in malaria. In one example, researchers are developing imaging technologies that they hope will allow them to visualize individual malaria molecules in live cells, and characterize the shape and deformation of infected host cells. And when recruiting, he looks for people with expertise in combinatorial chemistry, as well as in skills used to study nucleic acid–protein interactions and genome-wide responses of cells to drugs.

CHANGING RESEARCH CLIMATE

The funding climate indicates that these changes won't just be limited to Rathod's lab. Funding for malaria research by the US National Institutes of Health's National Institute of Allergy and Infectious Diseases (NIAID), for example, increased fivefold between financial years 1991 and 2001 (US\$10.4 million and \$57.8 million, respectively) and is estimated to reach \$72.1 million in 2003. The *P. falciparum* genome project was completed with funds from both public (NIAID and the US Department of Defense) and private (Burroughs Wellcome Fund and the Wellcome Trust) sectors.

The Seattle-based Bill and Melinda Gates Foundation has also supported research in this area, so far contributing more than \$125 million to malaria research. This includes \$50 million over ten years to the Malaria Vaccine Initiative at the Program for Appropriate Technology in Health (PATH) in Seattle and \$40 million over five years to the London School of Hygiene and Tropical Medicine in Britain for the

Pradipsinh Rathod welcomes scientists of all disciplines to malaria research.





S. STAMMERS/SPL creation of a new malaria centre.

Another promising new initiative, made possible by a \$100 million gift from an anonymous donor, was the creation in May 2001 of an institute dedicated solely to malaria research within the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. The Johns Hopkins Malaria Research Institute will bring together experts from a variety of disciplines — entomology, parasitology and bioinformatics — who are interested in revisiting issues from a basic-science point of view yet are squarely focused on the goal of developing new treatments and preventive measures for malaria.

The institute has already sparked “tremendous interest”, and not just from people in the field, says Nirbhay Kumar, professor of molecular microbiology and immunology at the new institute. Three new faculty have already been hired (two more offers are pending), and he expects to recruit five more over the next few years in the areas of bioinformatics, cell biology, immunology, structural biology and vector biology. There will also be fellowship money for training young students and postdoctoral fellows, as well as seed money for pilot projects that bring in people from elsewhere in the university who have never worked on malaria before.

The availability of the genome sequences “is going to open up tremendous opportunities for people to jump into an exciting and very challenging area”, says Kumar, who has been at Johns Hopkins since 1986 and has worked on malaria for 20 years. Originally from India,

Mosquito genomics takes off

Now that the sequence of *Anopheles gambiae*, the mosquito vector of the malaria parasite *Plasmodium falciparum*, is available, entomologist Peter Atkinson of the department of entomology at the University of California, Riverside, couldn't be happier. He works on mosquito-vector biology, developing genetic strategies for the control of insects that transmit malaria and other diseases.

The sequence gives researchers the ability to look through the whole genome for genes that are essential for

pathogen transmission through the mosquito, such as those involved in blood feeding in female mosquitoes.

When asked how his lab is coping with the torrent of sequence information flowing from the genome sequencing projects, he quips: “Well there are more computers than there used to be”. Atkinson's lab is made up principally of molecular biologists and geneticists and, when recruiting, he now looks for people with more computational skills than ever before. If they can also write code, that's a bonus.

D.G.

he has focused on the biology of malaria transmission and the development of better targets for transmission-blocking vaccines. He recalls a time when, as a postdoctoral fellow, he spent more than a year just trying to clone a gene — “a gene that I can get pretty much overnight” now that the sequence is available.

The field is “wide open”, he says. “The challenge is again upon scientists to make use of this information and learn from it and try to outsmart the parasite that has been outsmarting human endeavours so far.” ■

Diane Gershon is Assistant Editor, New Technology at *Nature Medicine*.

Nirbhay Kumar (left) of the new Johns Hopkins Malaria Research Institute sees great opportunities ahead.

The Bill and Melinda Gates Foundation
♦ <http://www.gatesfoundation.org>

Wellcome Trust

♦ <http://www.wellcome.ac.uk>

Burroughs Wellcome Fund

♦ <http://www.bwfund.org>

The Johns Hopkins Malaria Research Institute

♦ http://www.jhsph.edu/Dept/MMI/Malaria_Research_Institute/

London School of Hygiene and Tropical Medicine

♦ <http://www.lshtm.ac.uk>

Malaria Vaccine Initiative at PATH

♦ <http://www.malariavaccine.org>